반복측정 자료의 분석 고찰

서울대학교 보건대학원
김 호

1) 짝지은 t-검정 (시작하며)
2) 연속형 결과변수 (Proc Mixed)
3) 이산형 결과변수 (Proc Genmod)

Single Sample Analysis

- dataset: peppers

Peppers Dataset

Obs	angle
1	3
2	11
3	-7
4	2
5	3
6	8
7	-3
8	-2
9	13
10	4
11	7
12	-1
13	4
14	7
15	-1
16	4
17	12
18	-3
19	7
20	5
21	3
22	-1
23	9
24	-7
25	2
26	4
27	8
28	-2

```
proc means data=peppers mean std stderr t probt;
run;
```

Options

1. stderr: the standard error of the mean
2. t: $H_0 : \mu = 0$을 검정하는 t test
3. probt: the significance probability of the t test

The MEANS Procedure

분석 변수 : angle

평균값	표준편차	표준오차	t값	Pr > |t|
3.1785714	5.2988718	1.0013926	3.17	0.0037

Two Independent Samples

- dataset: bullets

Bullets Dataset

Obs	powder	velocity
1	1	27.3
2	1	28.1
3	1	27.4
4	1	27.7
5	1	28.0
6	1	28.1
7	1	27.4
8	1	27.1
9	2	28.3
10	2	27.9
11	2	28.1
12	2	28.3
13	2	27.9
14	2	27.6
15	2	28.5
16	2	27.9
17	2	28.4
18	2	27.7

```
proc ttest data=bullets;
var velocity; class powder;
run;
```

The TTEST Procedure

Statistics

Variable	powder	N	Lower CL Mean	Mean	Upper CL Mean	Lower CL Std Dev
velocity	1	8	27.309	27.638	27.966	0.2596
velocity	2	10	27.841	28.06	28.279	0.2106
velocity	Diff (1-2)		-0.771	-0.422	-0.074	0.2582

Statistics

Variable	powder	Std Dev	Upper CL Std Dev	Std Err	Minimum	Maximum
velocity	1	0.3926	0.799	0.1388	27.1	28.1
velocity	2	0.3062	0.5591	0.0968	27.6	28.5
velocity	Diff (1-2)	0.3467	0.5276	0.1644		

T-Tests

Variable	Method	Variances	DF	t Value	Pr > |t|
velocity	Pooled	Equal	16	-2.57	0.0206
velocity	Satterthwaite	Unequal	13.1	-2.50	0.0267

Equality of Variances

Variable	Method	Num DF	Den DF	F Value	Pr > F
velocity	Folded F	7	9	1.64	0.4782

```
data pulse;
input id pre post;
d=post-pre;
cards;
1 62 61
2 63 62
3 58 59
4 64 61
5 64 63
6 61 58
7 68 61
8 66 64
9 65 62
10 67 68
11 69 65
12 61 60
13 64 65
14 61 63
15 63 62
;
run;
```

```
data two;
set pulse;
x=0;
y=pre; output;
x=1;
y=post; output;

proc sort ;
by x ;
proc print noobs ;
var id x y; run;
proc ttest;
class x;
var y;

proc ttest data=pulse;
var d;
run;
```

id	x	y
1	0	62
2	0	63
3	0	58
4	0	64
5	0	64
6	0	61
7	0	68
8	0	66
9	0	65
10	0	67
11	0	69
12	0	61
13	0	64
14	0	61
15	0	63
1	1	61
2	1	62
3	1	59
4	1	61
5	1	63
6	1	58
7	1	61
8	1	64
9	1	62
10	1	68
11	1	65
12	1	60
13	1	65
14	1	63
15	1	62

The TTEST Procedure

Statistics

Variable	x	N	Lower CL Mean	Mean	Upper CL Mean	Lower CL Std Dev
y	0	15	62.092	63.733	65.374	2.1695
y	1	15	60.855	62.267	63.678	1.8659
y	Diff (1-2)		-0.601	**1.4667**	3.5338	2.1932

Statistics

Variable	x	Std Dev	Upper CL Std Dev	Std Err	Minimum	Maximum
y	0	2.9633	4.6734	0.7651	58	69
y	1	2.5486	4.0194	0.658	58	68
y	Diff (1-2)	2.7637	3.7378	**1.0092**		

T-Tests

Variable	Method	Variances	DF	t Value	Pr > |t|
y	Pooled	Equal	28	1.45	0.1572
y	Satterthwaite	Unequal	27.4	1.45	0.1575

Equality of Variances

Variable	Method	Num DF	Den DF	F Value	Pr > F
y	Folded F	14	14	1.35	0.5802

The TTEST Procedure

Statistics

Variable	N	Lower CL Mean	Mean	Upper CL Mean	Lower CL Std Dev	Std Dev
d	15	-2.755	**-1.467**	-0.179	1.7028	2.3258

Statistics

Variable	Upper CL Std Dev	Std Err	Minimum	Maximum
d	3.6681	**0.6005**	-7	2

T-Tests

Variable	DF	t Value	Pr > |t|
d	14	-2.44	0.0285

$E(Y_1-Y_2)= E(Y_1)-E(Y_2)$

$Var(Y_1-Y_2)= Var(Y_1)+Var(Y_2)-2\ Cov\ (Y_1,Y_2)$

$$Corr(Y_1,Y_2)=\frac{Cov(Y_1,Y_2)}{\sqrt{Var(Y_1)\cdot Var(Y_2)}}$$

- Paired t-test는 correlation을 자동으로 보정해줌
- Correlation between the same subject is very important !!
- 한 개체 안에서 3번 이상의 반복측정치가 있는 경우에는 짝지은 t-검정을 사용하기 곤란함
- 이 경우 correlation structure가 중요함

반복측정자료(repeated measures data)

- 하나의 관측단위로부터 두 번 이상의 측정을 통하여 얻어진 자료

- Ex 1) 신생아들의 체중을 한 달 간격으로 생후 일년간 측정, 기록
  2) 두 처방에서의 결과(예: 체중, 혈압 혹은 심장박동수 등)를 석 달마다 일년간에 걸쳐 관측 기록

- 두 가지의 중요한 요인

1) 시간: 환자내 효과(within-subjects factors)
2) 처리: 환자간 효과(between-subjects factors)

- 두 가지 관심 사항은

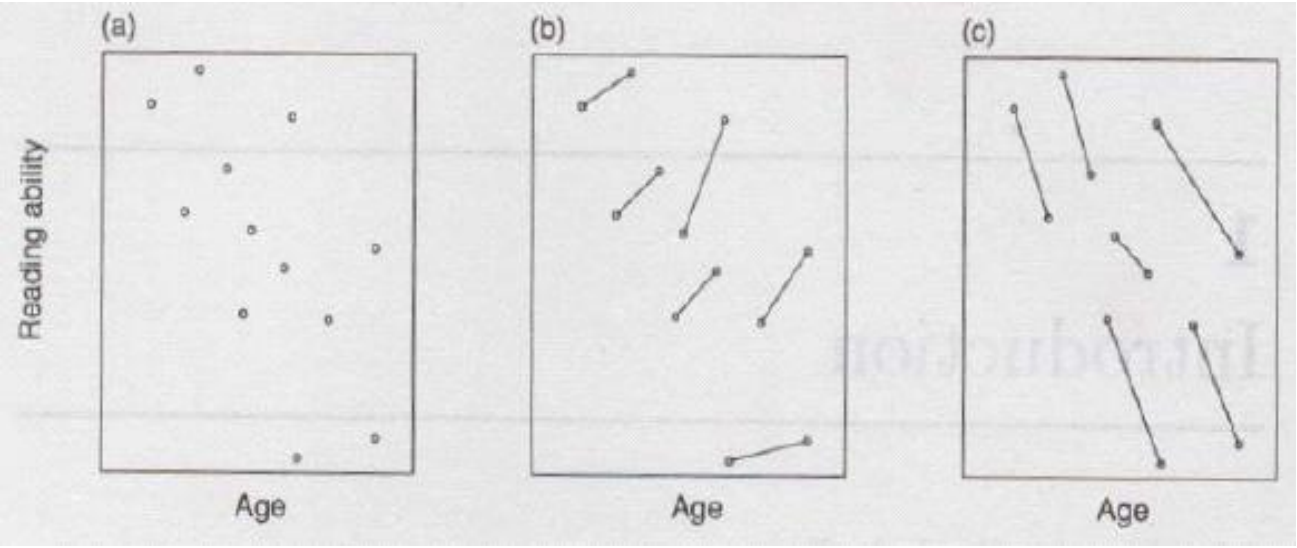
1) 각 처리의 평균값이 시간에 따라 어떻게 변하는지 ⟶ 시간의 주효과
2) 처리의 효과(처리간의 차이)가 시간에 따라 어떻게 변하는지 ⟶ 시간과 처리의 교호작용

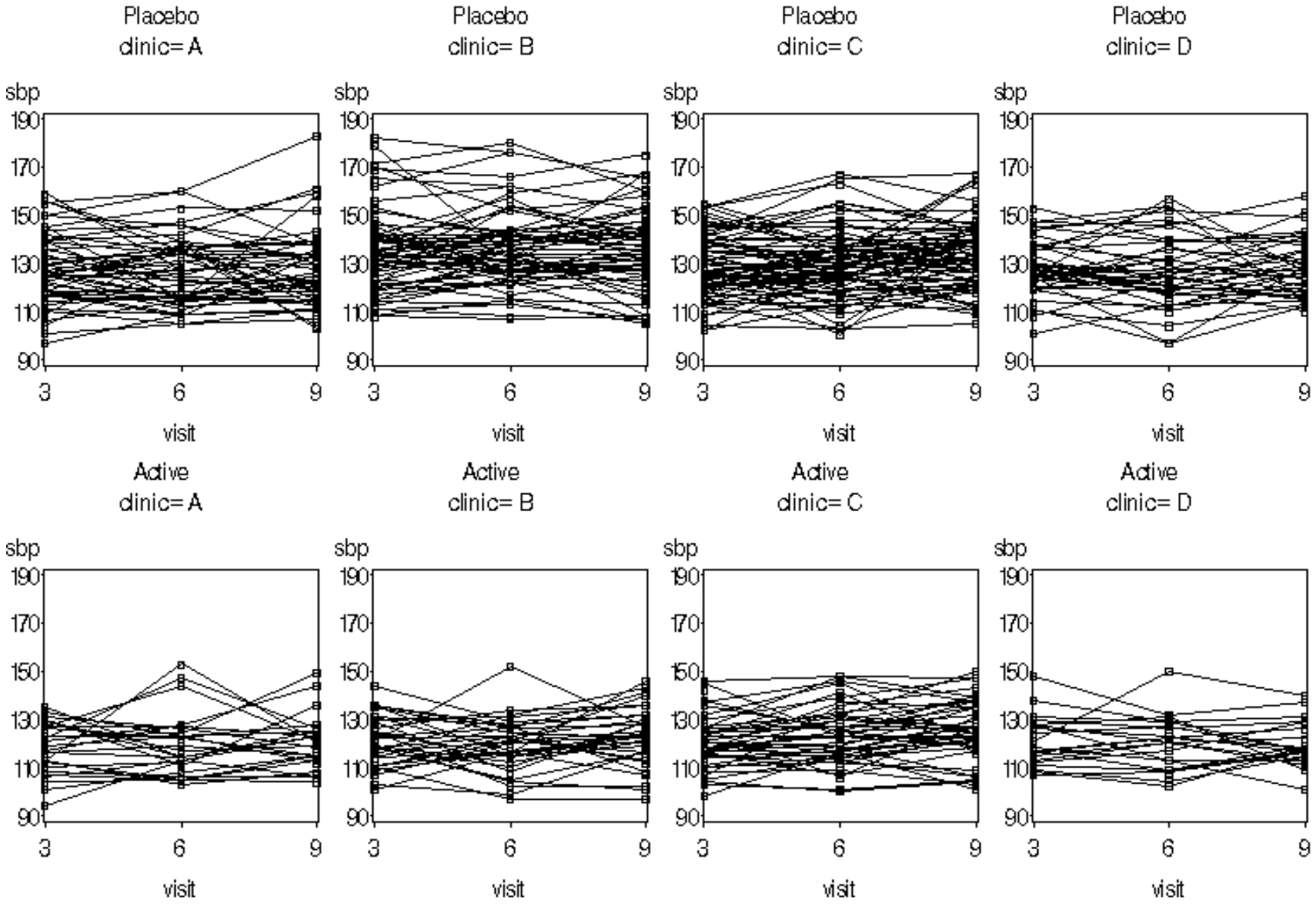
- 공분산구조 : 반복자료 분석이 일반적인 통계분석과 가장 다른 점
  - 일반적인 통계 선형모형에서는 각 관측치의 오차항은 독립적이라는 가정이 필수적
  - 반복측정 자료분석에서는
    - 각 각의 환자들은 독립적
    - 한 환자 안에서의 측정치, 즉 같은 환자의 다른 시점들에서 관측된 값들의 오차항 사이에는 상관관계가 존재한다고 가정
    - 한 환자 안의 관찰치들 간의 상관관계는 관측 시점의 간격에 따라 다르게 가정되는 것이 보통

  - 반복측정 자료 분석의 주요 목적은 시간에 따른 처리효과의 비교이지만, 모형의 구축 단계에서는 이 상관관계 구조의 설정에 가장 많은 노력을 기울임. 상관관계구조의 올바른 선택은 반복측정 자료 분석에서 가장 중요한 과정.
  - Example : BP data

- Treatment of Mild Hypertension Trial(TOMHS) by Neaton, et.al. (1993)

  - 경미한 고혈압 환자의 치료를 위한 확률화(randomized), 양쪽-맹검법(double-blind), 비교-처치군(placebo-controlled)을 이용한 임상
  - 연구의 주목적: 비교군에 비하여 처치군에서 약효를 확인할 수 있는가 하는 것이다.
  - <그림 2> 반복자료 분석의 시작 (그림 그리기)




Placebo
clinic= A
Placebo
clinic= B
Placebo
clinic= C
Placebo
clinic= D
Active
clinic= A
Active
clinic= B
Active
clinic= C
Active
clinic= D
sbp
190
170
150
130
110
90
3
6
9
visit

결측치에 대한 문제

ㅡ연구 초기에는 대부분의 환자에게서 자료

-연구가 진행될수록 제외(lost to follow-up or drop-out)되는 환자가 증가

- BP 데이터 3번의 follow-up 중 하나 이상에서 결측치 있는 환자는 13%

<표 2> Proc GLM 과 PROC MIXED 에서의 입력자료의 비교

PROC GLM	PROC MIXED
Id y1 y2 y3 x1 x2 x3	Id y1 x1 x2 x3 Id y2 x1 x2 x3 Id y3 x1 x2 x3
1 10 12 13 1 2 3 2 11 13 . 2 1 4	1 10 1 2 3 1 12 1 2 3 1 13 1 2 3 2 11 2 1 4 2 13 2 1 4 2 . 2 1 4

- 데이터 분석의 기본 목적:
  - 다른 모든 설명변수의 효과를 제어한 후의 SBP에 대한 TRT의 효과의 유의성을 보는 것
  - 동시에 같은 환자들의 관측치가 가지는 가능한 상관관계를 제어

- SBP값들은 다변량 정규분포를 따른다는 가정
  - 평균은 설명 변수들의 선형모형으로 표현
  - 비교적 간단한 구조의 분산-공분산

분산-공분산 행렬(variance-covariance Matrix)

- 단변량인 경우
- 다변량 (삼변량) 경우

$$Var(\varepsilon) = Var\begin{pmatrix}\varepsilon_1\\ \varepsilon_2\\ \varepsilon_3\end{pmatrix} = \begin{pmatrix}\text{var}(\varepsilon_1) & \text{cov}(\varepsilon_1,\varepsilon_2) & \text{cov}(\varepsilon_1,\varepsilon_3)\\ \text{cov}(\varepsilon_2,\varepsilon_1) & \text{var}(\varepsilon_2) & \text{cov}(\varepsilon_2,\varepsilon_3)\\ \text{cov}(\varepsilon_3,\varepsilon_1) & \text{cov}(\varepsilon_3,\varepsilon_2) & \text{var}(\varepsilon_3)\end{pmatrix} = \begin{pmatrix}\sigma_{11} & \sigma_{12} & \sigma_{13}\\ \sigma_{21} & \sigma_{22} & \sigma_{23}\\ \sigma_{31} & \sigma_{32} & \sigma_{33}\end{pmatrix}$$

- 혼합대칭(compound symmetry) 공분산 모형

$$Var(\varepsilon) = Var\begin{pmatrix}\varepsilon_1\\ \varepsilon_2\\ \varepsilon_3\end{pmatrix} = \begin{pmatrix}\sigma_1+\sigma^2 & \sigma_1 & \sigma_1\\ \sigma_1 & \sigma_1+\sigma^2 & \sigma_1\\ \sigma_1 & \sigma_1 & \sigma_1+\sigma^2\end{pmatrix}$$

- 일반선형모형 (독립인 오차항)

$$Var(\varepsilon) = Var\begin{pmatrix}\varepsilon_1\\ \varepsilon_2\\ \varepsilon_3\end{pmatrix} = \begin{pmatrix}\sigma^2 & 0 & 0\\ 0 & \sigma^2 & 0\\ 0 & 0 & \sigma^2\end{pmatrix}$$

고정효과(fixed effect)와 임의효과(random effect)

- 고정효과
  - 전통적인 선형모형에서 고려하는 효과
  - 하나의 고정된 값(모수)이고 우리는 적절한 통계적 방법을 통하여 이들을 추정
  - 연구자가 관심이 있는 효과의 수준이 일정하고 모든 수준에서 자료를 수집했다면

- 임의효과
  - 고정된 값이 아니고 분포를 가진 효과로서
  - 전체 분포에서 하나의 실현된 경우를 데이터를 통하여 관측
  - 효과의 분산을 추성
  - 효과의 수준이 다양하고 자료에서 관측하는 효과들은 전체 효과의 일부분으로 가정

- 고정효과의 예) 특정 처방에 대한 효과
  - 그 증상에 대한 처방이 다양하게 존재하지 않고
  - 연구자가 수집한 방법에만 관심이 있다면(예를 들어 두 가지 방법을 비교한다고 할 때 이 두 가지 처방이 여러 가지 가능한 처방 방법에서의 일부 샘플이라고 생각하지 않고 연구자가 이 두 가지 처방의 비교에만 관심이 있고, 연구의 결과를 이 두 가지 처방에 대한 비교로만 국한시켜도 무방)

- 임의효과의 예) 환자의 효과
  - 다양하게 존재하고
  - 연구자가 모든 환자들의 효과를 수집할 수도 없고
  - 자료에서의 환자효과는 전체 연구집단에서의 임의 추출(random sampling)된 것이라고 가정해도 무방

❑ 두 가지 분석 방법

1) 최초 관측치(baseline observation)를 설명변수로 하고 나머지 관측치를 종속변수
2) 최초 관측치와 각 시점에서의 관측치의 차이를 종속변수로

- 결국 동등한 모형으로
- 해석하고자 하는 방향에 따라 선택
- 여기에서는 전자의 방법을 사용
- SBP의 최초방문 시에 비한 SBP의 증가분이 아닌 최초방문 SBP 값으로 보정한 후의 각각 방문 시점에서의 SBP값에 관심이 있게 된다.

❑ 변수설명

Sbp : 혈압 (반응변수)

Sbpbl : 최초 혈압

trt : 처리 (1=Active, 2=Placebo)

Visit : 재방문 개월 (3,6,9 class 변수)

Complier : 순응도 (1 or 0)

clinic : 병원 (A, B, C, or D)

Stratum : 고혈압 약 복용 경험여부 (1 or 0)

Visitln:연속변수로서의 재방문 개월(실제 값은 visit과 동일)

<모형 1>일반 선형 모형(General Linear Model)

- 서로 독립이고 동일한 분포를 가지는(iid) 오차항 가정

```
proc mixed data=bp;
   class trt visit complier clinic stratum;
   model sbp = sbpbl trt
      visit trt*visit
      complier trt*complier
      clinic trt*clinic
      stratum trt*stratum;

run;
```

- PROC MIXED는 혼합효과 모형을 실행하게 하고 DATA=BP는 사용할 데이터의 이름을 나타낸다. CLASS문에서는 비연속 변수를 설정해 주고 MODEL문에서 고려되는 설명변수들을 설정해주고 있다. #cov parameter = 1

<모형 2>혼합대칭 공분산(compound Symmetry)모형

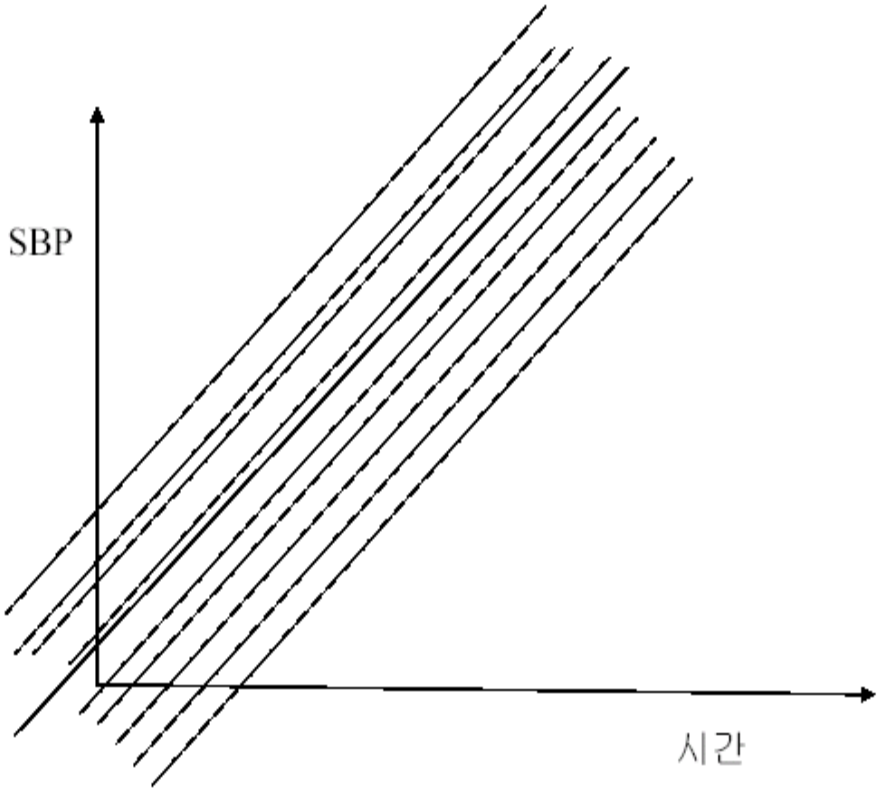
```
proc mixed data=bp;
   class trt visit complier clinic stratum person;
   model sbp = sbpbl trt
      visit trt*visit
      complier trt*complier
      clinic trt*clinic
      stratum trt*stratum;
   repeated visit / type=cs sub=person;
run;
```

- Type=CS : 대각선은 공통의 값을 가지고 비대각의 값은 또 다른 값을 가지는 형태, #parameter=2

<모형 2>와 동등한 임의 절편모형

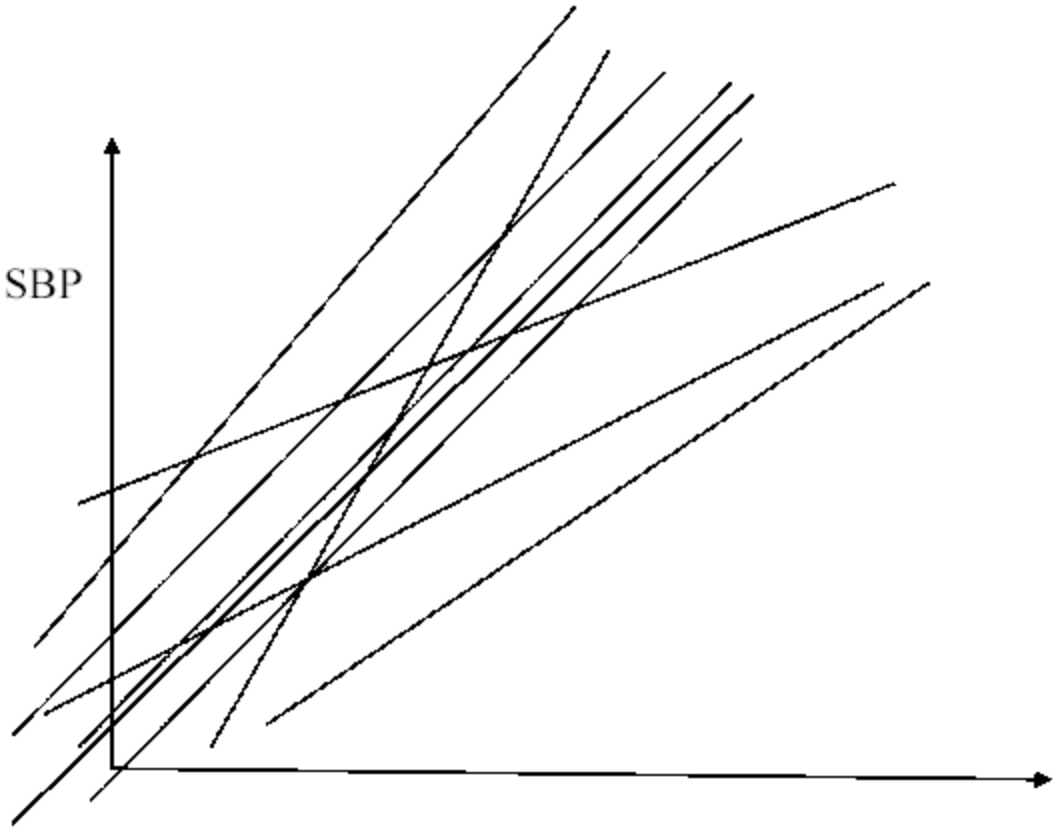
```
proc mixed data=bp;
    class trt visit complier clinic stratum person;
    model sbp = sbpbl trt
        visit trt*visit
        complier trt*complie
        clinic trt*clinic
        stratum trt*stratum;
    random int / sub=person
run;
```

#parameter=2



<모형 3>임의계수(Random Coefficients)모형

```
proc mixed data=bp;
    class trt visit complier clinic stratum person;
    model sbp = sbpbl trt
        visit trt*visit
        complier trt*complier
        clinic trt*clinic
        stratum trt*stratur
    random int visitlin /
   type=un sub=person;
run;

#parameter = 4
```



<모형 4> 비구조화(Unstructured) 분산 모형

```
proc mixed data=bp;
   class trt visit complier clinic stratum person;
   model sbp = sbpbl trt
      visit trt*visit
      complier trt*complier
      clinic trt*clinic
      stratum trt*stratum;
   repeated visit / type=un sub=person r rcorr;
run;

# patameter = 6 (분산 3개, 공분산 3개)
```

<모형 5> 이질성(heterogeneous) 일반 선형 모형

- 두 군간에 분산이 다른 것을 모형화

```
proc mixed data=bp;
    class trt visit complier clinic stratum;
    model sbp = sbpbl trt
        visit trt*visit
        complier trt*complier
        clinic trt*clinic
        stratum trt*stratum;
    repeated / sub=person group=trt;
run;

# parameter = 2
```

<모형 6> 이질성 혼합대칭 모형

```
proc mixed data=bp;
   class trt visit complier clinic stratum person;
   model sbp = sbpbl trt
      visit trt*visit
      complier trt*complier
      clinic trt*clinic
      stratum trt*stratum;
   repeated visit / type=cs sub=person group=trt;
run;
```

parameters = 4, 비교군에서 2개, 처리군에서 2개

<모형 7> 이질성 임의계수 모형

```
proc mixed data=bp;
   class trt visit complier clinic stratum person;
   model sbp = sbpbl trt
      visit trt*visit
      complier trt*complier
      clinic trt*clinic
      stratum trt*stratum / ddfm=bw;
   random int visitlin / type=un sub=person group=trt;
   repeated / sub=person group=trt;
run;
```

parameters = 8, 각각의 처치군에서 4개씩 8개

<모형 8> 이질성 비구조화 분산 모형

```
proc mixed data=bp;
   class trt visit complier clinic stratum person;
   model sbp = sbpbl trt
      visit trt*visit
      complier trt*complier
      clinic trt*clinic
      stratum trt*stratum;
   repeated visit / type=un sub=person group=trt;
run;
```

parameters = 12 각각의 처치군에서 6개씩 12개

<표 3> BP 데이터를 분석하기 위한 모형들

번호	모형	분산-공분산 행렬의 모수 개수	BIC
1	일반 선형화	1	-4300.5
2	혼합대칭	2	-3957.6
3	임의 계수	4	-3962.2
4	비구조화	6	-3968.2
5	이질성 일반 선형화	2	-4027.1
6	이질성 혼합대칭	4	-3957.1
7	이질성 임의계수	8	-3968.3
8	이질성 비구조화	12	-3980.5

❑ 모형을 적합시킨 후에는 잔차 검사한다 (잔차분석)

```
proc mixed data=bp;
   class trt visit complier clinic stratum person;
   model sbp = sbpbl trt
      visit trt*visit
      complier trt*complier
      clinic trt*clinic
      stratum trt*stratum / outp=p;
   repeated visit / type=cs sub=person group=trt ;
   make 'predicted' out=p noprint;
   id trt visit clinic person;
run;
```

Covariance Parameter Estimates

Cov Parm	Subject	Group	Estimate	
Variance	person	trt 1	69.0629	$=\widehat{\sigma}^2$
CS	person	trt 1	34.9670	$=\widehat{\sigma}_1$
Variance	person	trt 2	87.2927	$=\widehat{\sigma}^2$
CS	person	trt 2	71.7782	$=\widehat{\sigma}_1$

Fit Statistics

-2 Res Log Likelihood	7886.5
AIC (smaller is better)	7894.5
AICC (smaller is better)	7894.6
BIC (smaller is better)	7910.1

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
visit	2	682	4.90	0.0077
trt*visit	2	682	0.45	0.6393
complier	1	345	0.47	0.4914
trt*complier	1	345	0.05	0.8263
clinic	3	345	6.25	0.0004
trt*clinic	3	345	3.76	0.0112
stratum	1	345	0.06	0.8021
trt*stratum	1	345	1.86	0.1736

```
proc mixed data=bp;
   class trt visit complier clinic stratum person;
   model sbp = sbpbl trt visit clinic trt*clinic/s;
   repeated visit / type=cs sub=person group=trt;
   lsmeans trt*clinic / cl;
run;
```

Least Squares Means

Effect	clinic	trt	Estimate	Standard Error	DF	t Value	Pr > |t|
trt*clinic	A	1	121.30	1.4817	349	81.87	<.0001
trt*clinic	B	1	123.21	1.3044	349	94.46	<.0001
trt*clinic	C	1	122.46	1.1529	349	106.21	<.0001
trt*clinic	D	1	122.38	1.6317	349	75.00	<.0001
trt*clinic	A	2	126.79	1.4116	349	89.82	<.0001
trt*clinic	B	2	136.02	1.2604	349	107.92	<.0001
trt*clinic	C	2	129.45	1.1777	349	109.92	<.0001
trt*clinic	D	2	127.90	1.5799	349	80.95	<.0001

❑ TRT효과 (CLINIC과의 교호작용에 유의)

CLINIC= A에서는 126.79-121.30=5.49,

B에서는 136.02-123.21=12.81,

C에서는 129.45-122.46=6.99,

D에서는 127.90-122.38=5.52

```
lsmeans trt / diff cl e;
lsmeans trt / at sbpbl=170 cl;
```

Least Squares Means

Effect	trt	sbpbl	Estimate	Standard Error	DF	t Value	Pr > |t|
trt	1	140.02	122.34	0.7051	349	173.51	<.0001
trt	2	140.02	130.04	0.6834	349	190.29	<.0001
trt	1	170.00	138.64	1.4373	349	96.46	<.0001
trt	2	170.00	146.35	1.3221	349	110.69	<.0001

Differences of Least Squares Means

Effect	trt	_trt	sbpbl	Estimate	Standard Error	DF	t Value	Pr > |t|
trt	1	2	140.02	-7.7021	0.9848	349	-7.82	<.0001

- 이산형 결과변수의 경우 : GEE(Generalized Estimating Equations)
- Working correlation structures
- Exchangeable : all correlations are the same

- Stationary m-dependent

$$Corr(Y_{ij}, Y_{ij+1}) = \rho_1$$
$$Corr(Y_{ij}, Y_{ij+2}) = \rho_2$$
$$\cdots$$
$$Corr(Y_{ij}, Y_{ij+m}) = \rho_m$$
$$Corr(Y_{ij}, Y_{ij+m'}) = 0, \quad if \quad m' > m$$

$$n_i = 4, R_i(\rho_1) = \begin{bmatrix} 1 & \rho_1 & 0 & 0 \\ \rho_1 & 1 & \rho_1 & 0 \\ 0 & \rho_1 & 1 & \rho_1 \\ 0 & 0 & \rho_1 & 1 \end{bmatrix}$$

- Example 1. Stationary 1-dependant

- Example 2. Stationary 2-dependent

$$n_i = 5, R_i(\rho_1, \rho_2) = \begin{bmatrix} 1 & \rho_1 & \rho_2 & 0 & 0 \\ \rho_1 & 1 & \rho_1 & \rho_2 & 0 \\ \rho_2 & \rho_1 & 1 & \rho_1 & \rho_2 \\ 0 & \rho_2 & \rho_1 & 1 & \rho_1 \\ 0 & 0 & \rho_2 & \rho_1 & 1 \end{bmatrix}$$

- Autoregressive (AR-1)

$$Corr(Y_{ij}, Y_{ij'}) = \rho^{|t_{ij} - t_{ij'}|}$$

- Example.

$$n_i = 5, R_i(\rho) = \begin{bmatrix} 1 & \rho & \rho^2 & \rho^3 & \rho^4 \\ \rho & 1 & \rho & \rho^2 & \rho^3 \\ \rho^2 & \rho & 1 & \rho & \rho^2 \\ \rho^3 & \rho^2 & \rho & 1 & \rho \\ \rho^4 & \rho^3 & \rho^2 & \rho & 1 \end{bmatrix}$$

- Arbitrary correlation
  - No restriction on $R(\alpha)$ so n(n+ 1)/2 elements

$$R_i(\alpha) = \begin{bmatrix} 1 & \rho_{12} & \rho_{13} & \rho_{14} \\ \rho_{12} & 1 & \rho_{23} & \rho_{24} \\ \rho_{13} & \rho_{23} & 1 & \rho_{34} \\ \rho_{14} & \rho_{24} & \rho_{34} & 1 \end{bmatrix}$$

- SAS options
  - AR|AR(1) : Autoregressive(1)
  - EXCH|CS : exchangeable
  - IND : independent
  - MDEP : m-dependent
  - UNSTR|UN : unrestricted (arbitrary)
  - USER|FIXED : fixed, user-specified correlation matrix

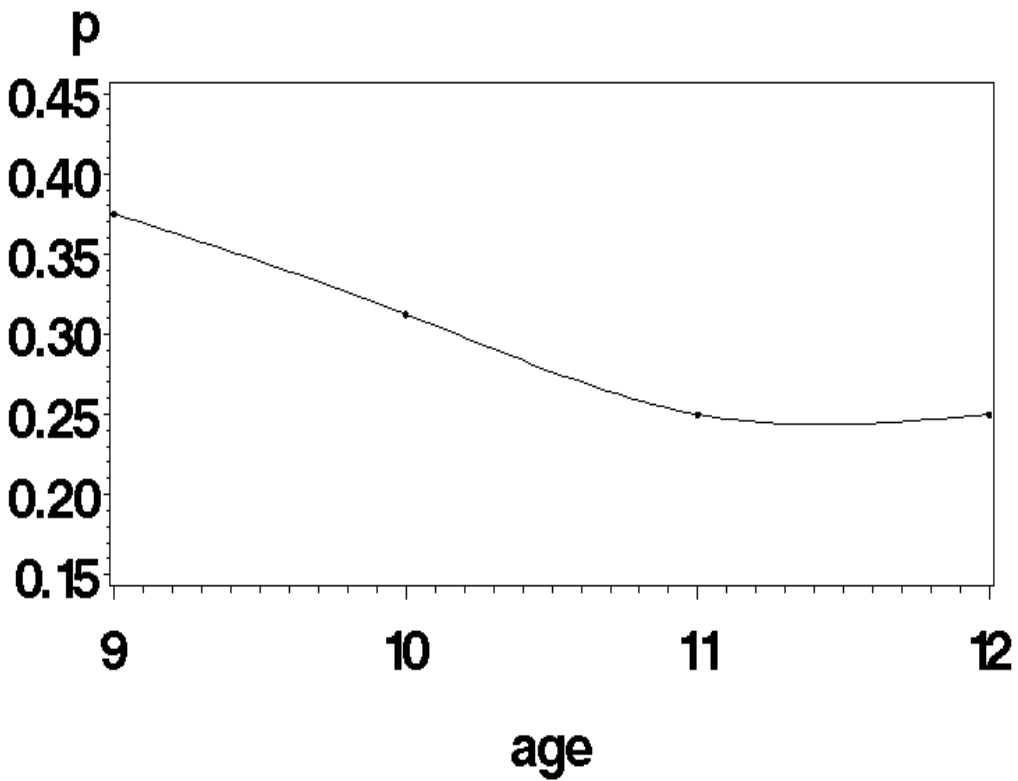
```
Type=user( 1.0 .9 .8 .6
           .9 1.0 .9 .8
           .8 .9 1.0 .9
           .6 .8 .9 1.0)
```

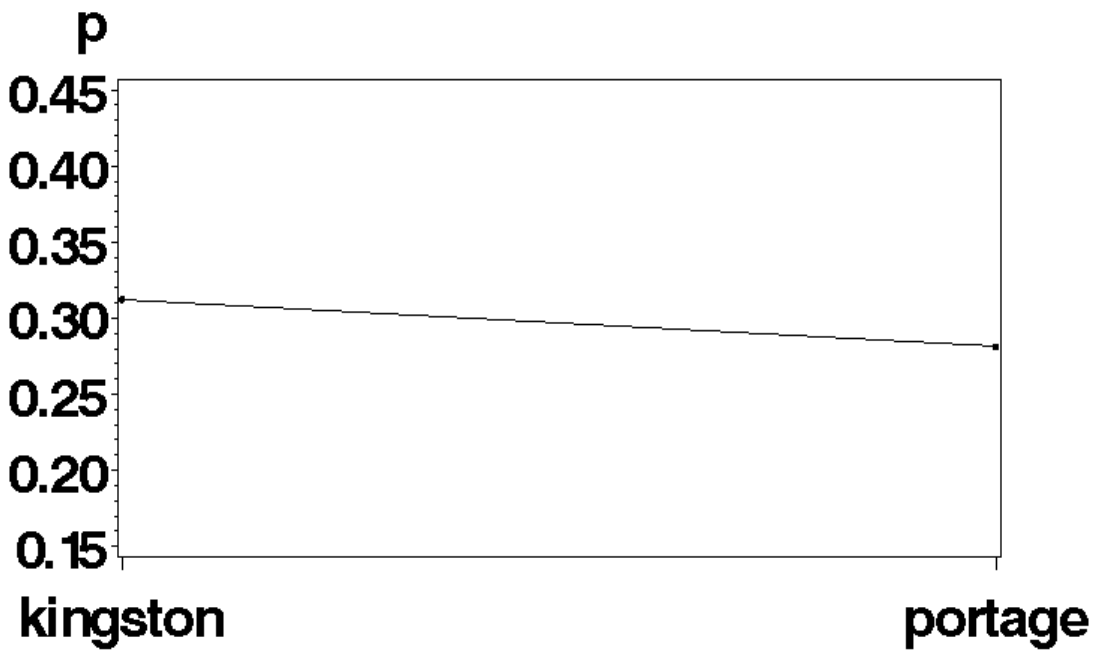
- Example : Six Cities Study of health effects on pollution

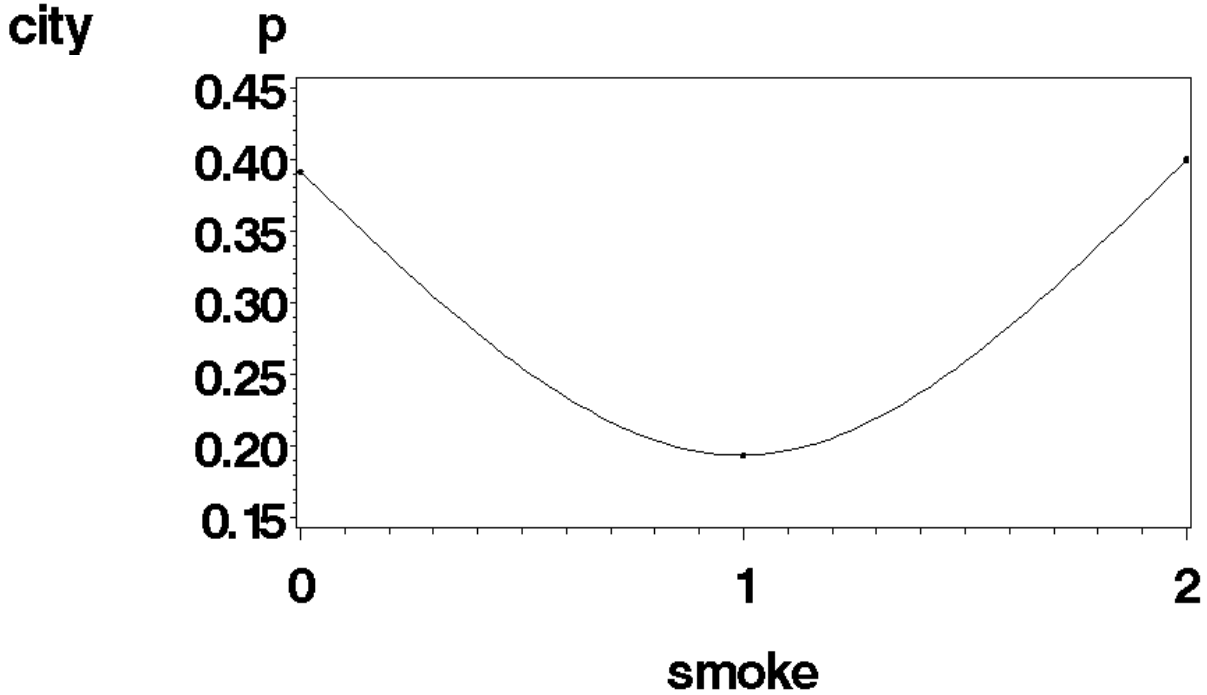
- Response : Wheezing status of sixteen children at ages 9, 10, 11, and 12.
- Explanatory vars : city of residence, age maternal smoking status at the particular age (time varying covariate)

```
data six;
   input case city$ @;
   do i=1 to 4;
        input age smoke wheeze @@;
        output;
   end;
   datalines;
 1 portage  9 0 1  10 0 1  11 0 1  12 0 0
 2 kingston 9 1 1  10 2 1  11 2 0  12 2 0
 3 kingston 9 0 1  10 0 0  11 1 0  12 1 0
 4 portage  9 0 0  10 0 1  11 0 1  12 1 0
 5 kingston 9 0 0  10 1 0  11 1 0  12 1 0
 6 portage  9 0 0  10 1 0  11 1 0  12 1 0
 7 kingston 9 1 0  10 1 0  11 0 0  12 0 0
 8 portage  9 1 0  10 1 0  11 1 0  12 2 0
 9 portage  9 2 1  10 2 0  11 1 0  12 1 0
10 kingston 9 0 0  10 0 0  11 0 0  12 1 0
11 kingston 9 1 1  10 0 0  11 0 1  12 0 1
12 portage  9 1 0  10 0 0  11 0 0  12 0 0
13 kingston 9 1 0  10 0 1  11 1 1  12 1 1
14 portage  9 1 0  10 2 0  11 1 0  12 2 1
15 kingston 9 1 0  10 1 0  11 1 0  12 2 1
16 portage  9 1 1  10 1 1  11 2 0  12 1 0
;
goptions gunit=pct border ftext=swissb htext=6;
```

```
%macro p(var) ;

proc sort data=six out=se;
by &var;

proc means data=se mean;
var wheeze;
by &var;
output out=s mean=p;

proc gplot;
symbol1 i=spline v=dot;
plot p*&var/vaxis=0.15 to 0.45 by 0.05;
run;

%mend p;

%p(age);
%p(city);
%p(smoke);
```




p
0.45
0.40
0.35
0.30
0.25
0.20
0.15
kingston
portage


city
p
0.45
0.40
0.35
0.30
0.25
0.20
0.15
0
1
2
smoke

```
proc genmod data=six;
   class case city smoke;
   model wheeze=city age smoke / dist=bin;
   repeated subject=case / type=exch covb corrw;
run;
```

The GENMOD Procedure

Model Information

Data Set	WORK.SIX
Distribution	Binomial
Link Function	Logit
Dependent Variable	wheeze
Observations Used	64
Probability Modeled	Pr(wheeze = 0)

Class Level Information

Class	Levels	Values
case	16	1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16
city	2	kingston portage
smoke	3	0 1 2

Response Profile

Ordered Level	Ordered Value	Count
1	0	45
2	1	19

Parameter Information

Parameter	Effect	city	smoke
Prm1	Intercept		
Prm2	city	kingston	
Prm3	city	portage	
Prm4	age		
Prm5	smoke		0
Prm6	smoke		1
Prm7	smoke		2

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	59	73.6976	1.2491
Scaled Deviance	59	73.6976	1.2491
Pearson Chi-Square	59	62.8302	1.0649
Scaled Pearson X2	59	62.8302	1.0649
Log Likelihood		-36.8488	

Algorithm converged.

Analysis Of Initial Parameter Estimates

Parameter		DF	Estimate	Standard Error	Wald 95% Confidence Limits	
Intercept		1	-2.1841	2.9166	-7.9006	3.5323
city	kingston	1	-0.2105	0.5695	-1.3266	0.9056
city	portage	0	0.0000	0.0000	0.0000	0.0000
age		1	0.2459	0.2619	-0.2674	0.7592
smoke	0	1	0.2003	0.7982	-1.3641	1.7647
smoke	1	1	1.1712	0.8143	-0.4249	2.7673
smoke	2	0	0.0000	0.0000	0.0000	0.0000
Scale		0	1.0000	0.0000	1.0000	1.0000

Analysis Of Initial Parameter Estimates

Parameter		Chi-Square	Pr > ChiSq
Intercept		0.56	0.4539
city	kingston	0.14	0.7116
city	portage	.	.
age		0.88	0.3478
smoke	0	0.06	0.8018
smoke	1	2.07	0.1504
smoke	2	.	.
Scale			

노트: The scale parameter was held fixed.

GEE Model Information

Correlation Structure	Exchangeable
Subject Effect	case (16 levels)
Number of Clusters	16
Correlation Matrix Dimension	4
Maximum Cluster Size	4
Minimum Cluster Size	4

Covariance Matrix (Model-Based)

	Prm1	Prm2	Prm4	Prm5	Prm6
Prm1	7.17537	-0.14111	-0.60885	-0.89483	-0.80310
Prm2	-0.14111	0.49270	-0.006190	-0.06016	-0.05615
Prm4	-0.60885	-0.006190	0.05604	0.04212	0.03830
Prm5	-0.89483	-0.06016	0.04212	0.72117	0.47716
Prm6	-0.80310	-0.05615	0.03830	0.47716	0.62375

Covariance Matrix (Empirical)

	Prm1	Prm2	Prm4	Prm5	Prm6
Prm1	7.96902	-0.86222	-0.76067	0.22448	0.35709
Prm2	-0.86222	0.45438	0.06677	-0.07237	-0.02958
Prm4	-0.76067	0.06677	0.07485	-0.03548	-0.06026
Prm5	0.22448	-0.07237	-0.03548	0.40782	0.40360
Prm6	0.35709	-0.02958	-0.06026	0.40360	0.64221

Algorithm converged.

Working Correlation Matrix

	Col1	Col2	Col3	Col4
Row1	1.0000	0.1837	0.1837	0.1837
Row2	0.1837	1.0000	0.1837	0.1837
Row3	0.1837	0.1837	1.0000	0.1837
Row4	0.1837	0.1837	0.1837	1.0000

Analysis Of GEE Parameter Estimates
Empirical Standard Error Estimates

Parameter		Estimate	Standard Error	95% Confidence Limits		Z	Pr >|Z|
Intercept		-2.1597	2.8229	-7.6926	3.3731	-0.77	0.4442
city	kingston	-0.1605	0.6741	-1.4817	1.1607	-0.24	0.8118
city	portage	0.0000	0.0000	0.0000	0.0000	.	.
age		0.2444	0.2736	-0.2918	0.7806	0.89	0.3716
smoke	0	0.2163	0.6386	-1.0353	1.4680	0.34	0.7348
smoke	1	1.0680	0.8014	-0.5027	2.6387	1.33	0.1826
smoke	2	0.0000	0.0000	0.0000	0.0000	.	.

Summary

- 반복측정 자료 분석은 한 개체 안에서 반복적으로 측정된 자료간에는 상관이 있다는 사실에 부합하는 분석 방법이다.
- 독립성을 가정한 일반 선형모형(일반 회귀분석이나 분산분석)을 사용하는 것은 가정에 어긋나는 모형을 사용하는 것이다.
- 연구의 목적이 상관관계의 분석이 아닌 약효의 차이에 대한 것이라도 반복자료 분석 기법을 이용해야만 상관성을 보정한 후의 효과를 얻을 수 있다.
- 올바른 약효를 추정하기 위해서는 올바른 공분산 구조를 선택해야 한다.
- 올바른 약효를 추정하기 위해서는 교호작용의 선택 등 변수 선택에 유의해야 한다.

김호 “반복측정 자료 분석의 고찰: SAS PROC MIXED를 이용하여” 한국보건통계학회지 1999, 24, pp.7-15.