



제 6 강. 반복측정 자료분석

이재원 교수
고려대학교 통계학과

주요내용

1. 반복측정 자료분석의 개념
2. 반복요인이 하나인 경우에 대한 반복측정
자료분석
3. 반복요인이 둘 이상인 경우에 대한 반복
측정 자료분석
4. R 프로그램을 이용한 분석

반복측정 자료분석

◆ 반복측정(Repeated measurements)

: 동일한 실험 단위에 대해 여러 번의 측정하는 것

예) 임상시험에서 동일한 환자가 병원을 여러 차례 반복적으로 방문하여 매 방문 때마다 반응치를 측정하는 경우



반복요인이 하나인 경우에 대한 반복측정 자료분석

<예제 5.3>

- ◆ <표 5.6>은 고혈압 환자 66명을 대상으로 32명에게는 기존의 혈압 강하제(C로 표기)를 투여하고 나머지 34명에게는 새로운 혈압강하제(E로 표기)를 투여하여, 각각 3회에 걸쳐 혈압을 측정한 자료이다.
즉, 첫 방문 시 약을 투여 받기 전 환자의 혈압과 4주 후, 8주 후의 혈압을 측정하였다.



<예제 5.3>



◆ 반복측정 자료분석에 대한 목표

1. 기존약과 새로운 혈압강하제 간의 유의한 차이가 있는가?
2. 방문시간에 따라 혈압에 차이가 있는가?
3. 약과 방문 시간 간에 교호작용이 있는가?

반복요인이 하나인 경우

< 표 5.7 >

		시간			
처리	개체	1	2	...	K
1	1	y_{111}	y_{112}	:	y_{11K}
	:	:	:	:	:
	n_1	$y_{1n_1 1}$	$y_{1n_1 2}$:	$y_{1n_1 K}$
2	1	y_{211}	y_{212}	:	y_{21K}
	:	:	:	:	:
	n_2	$y_{2n_2 1}$	$y_{2n_2 2}$:	$y_{2n_2 K}$
:	:	:	:	:	:
I	1	y_{I11}	y_{I12}	:	y_{I1K}
	:	:	:	:	:
	n_I	$y_{In_I 1}$	$y_{In_I 2}$:	$y_{In_I K}$

반복요인이 하나인 경우

▶ 반복요인이 하나인 반복측정 자료의 형태

- I : 처리의 수
- n_1, \dots, n_I : 각 처리에 할당된 개체수
- 첨자 i : 처리, 첨자 j : 개체, 첨자 k : 시간
- Y_{ijk} : i 번째 처리에서 매 k 시점마다 측정한 j 번째 반응 측정치

통계적 모형

$$y_{ijk} = \mu + \alpha_i + \beta_{ij} + \gamma_k + \delta_{ik} + \varepsilon_{ijk}$$

식(5.16)

$$(i = 1, \dots, I; \ j = 1, \dots, n_i; \ k = 1, \dots, K, \ \varepsilon_{ijk} \sim iid \ N(0, \sigma^2))$$

반복측정자료분석 모형설명

- μ : 반응측정치, y_{ijk} 의 전체평균을 나타내는 모수
- α_i : i 번째 처리효과(Treatment effect)
- β_{ij} : i 번째 처리내에서 개체 j 의 효과(Subject effect)
- γ_k : 시간 k 의 효과(Time effect)
- δ_{ik} : i 번째 처리와 k 번째 시간의 교호작용 효과(Interaction effect)
- ε_{ij} : 평균이 0이며 분산이 σ^2 인 독립적인 정규분포를 따른다고 가정

오차항의 구형성

- ◆ 복합대칭성(compound symmetry) 또는 구형성 (sphericity)

: 각 반복측정치들 간의 상관관계가 모두 동일함을 의미.

$$COV(\varepsilon)_{K \times K} = \sigma^2 \begin{pmatrix} 1 & \rho & \cdots & \rho \\ \rho & 1 & \cdots & \rho \\ \vdots & & \ddots & \\ \rho & \cdots & \rho & 1 \end{pmatrix}$$

즉, 모든 시간($k = 1, \dots, K$)에서의 분산은 σ^2 로

같고 공분산도 모든 시간에 대해 $\sigma^2 \rho$ 로 같음

오차항의 구형성

- ▶ 복합대칭성(compound symmetry) 또는 구형성 (sphericity):
반복측정 자료분석을 일변량으로 분석하기 위한 조건
- ▶ 구형성 가정 만족 → 일변량 분석
다변량 분석도 가능하지만 보수적
- ▶ 구형성 가정 불만족 → 다변량 분석
일변량 분석(자유도 수정)도 가능

귀무가설

1. $H_{01} : \alpha_1 = \alpha_2 = \dots = \alpha_I = 0$

-처리 간의 반응측정치의 차이가 없음

2. $H_{02} : \gamma_1 = \gamma_2 = \dots = \gamma_K = 0$

-시간 간의 반응측정치의 차이가 없음

3. $H_{03} : \delta_{11} = \delta_{12} = \dots = \delta_{IK} = 0$

-처리와 시간 간의 교호작용이 존재하지 않음

분산분석표

<표 5.8> 구형성을 만족하는 자료의 분산분석표

요인	자유도	제곱합	평균제곱합	F
처리	$I - 1$	$SSI = K \sum_{i=1}^I n_i (\bar{y}_{i..} - \bar{y}_{...})^2$	$SSI / (I - 1)$	F_1
개체(처리)	$N - I$	$SSJ = K \sum_{i=1}^I \sum_{j=1}^{n_i} (\bar{y}_{ij.} - \bar{y}_{i..})^2$	$SSJ / (N - I)$	
시간	$K - 1$	$SSK = N \sum_{k=1}^K (\bar{y}_{..k} - \bar{y}_{...})^2$	$SSK / (K - 1)$	F_2
처리*시간	$(I - 1)(K - 1)$	$SS_{I \times K} = \sum_{k=1}^K \sum_{i=1}^I n_i (\bar{y}_{i.k} - \bar{y}_{i..} - \bar{y}_{..k} + \bar{y}_{...})^2$	$SS_{I \times K} / (I - 1)(K - 1)$	F_3
오차	$(N - I)(K - 1)$	$SSE = \sum_{k=1}^K \sum_{i=1}^I \sum_{j=1}^{n_i} (y_{ijk} - \bar{y}_{ij.} - \bar{y}_{i.k} + \bar{y}_{i..})^2$	$SSE / (N - I)(K - 1)$	
계	$NK - 1$	$SST = \sum_{k=1}^K \sum_{i=1}^I \sum_{j=1}^{n_i} (y_{ijk} - \bar{y}_{...})^2$		

각 모수의 검정

◆ '처리 간 차이가 없다'는 귀무가설 H_{01} 검정

$$F_1 = \frac{SSI / (I - 1)}{SSJ / (N - I)} \sim F(I - 1, N - I) \quad (5.18)$$

◆ '시간의 효과가 없다'는 귀무가설 H_{02} 검정

$$F_2 = \frac{SSK / (K - 1)}{SSE / (N - I)(K - 1)} \sim F(K - 1, (N - I)(K - 1)) \quad (5.19)$$

◆ '처리와 시간의 교호작용이 없다'는 H_{03} 검정

$$F_3 = \frac{SS_{I \times K} / (I - 1)(K - 1)}{SSE / (N - I)(K - 1)} \sim F((I - 1)(K - 1), (N - I)(K - 1)) \quad (5.20)$$



<예제 5.3>

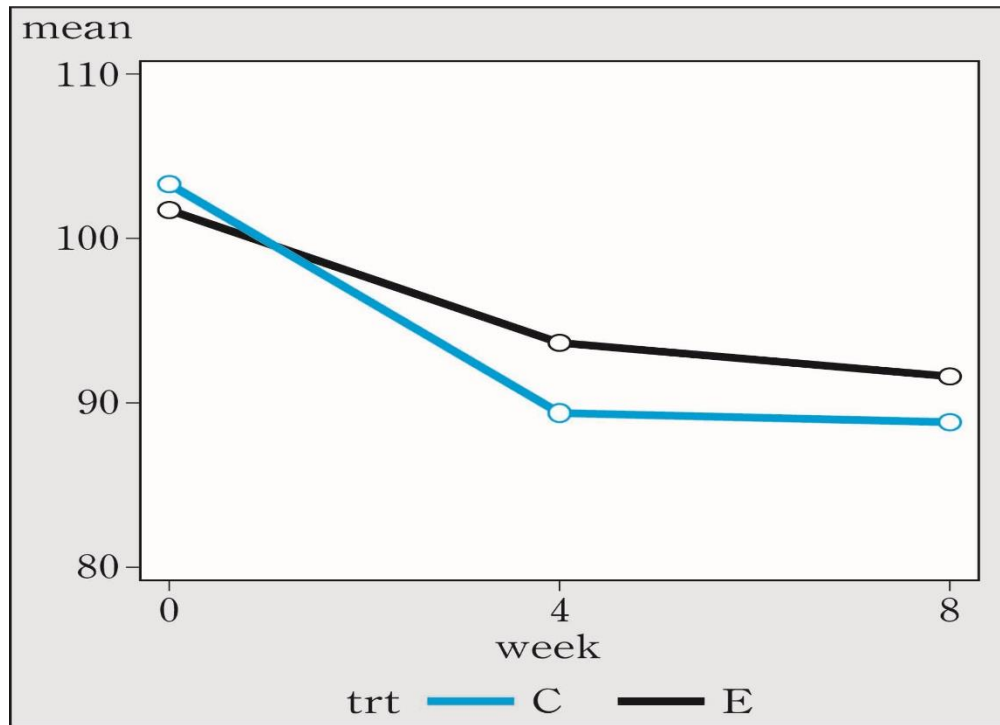


◆ 반복측정 자료분석에 대한 목표

1. 기존약과 새로운 혈압강하제 간의 유의한 차이가 있는가?
2. 방문시간에 따라 혈압에 차이가 있는가?
3. 약과 방문 시간 간에 교호작용이 있는가?

시간에 따른 평균 혈압

<그림 5.6>



R Program

```
library(car)
options(contrasts=c("contr.sum", "contr.poly"))
고혈압자료=read.table("고혈압자료.txt",header=T)

# 반복측정 자료분석
시간 = ordered(colnames(고혈압자료)[-4])
데이터 = data.frame(시간)
반복측정1.1 = lm(cbind(혈압.0주,혈압.4주,혈압.8주) ~ 혈압강하제,
                 data=고혈압자료)
반복측정1.2 = Anova(반복측정1.1, idata=데이터, idesign=~ 시간, type ="3")
```

- **ordered 함수** : 순서화된 인자(factor)로 만들어 줌.
- **Anova 함수(Car 라이브러리)** : 함수의 인수(argument) 중 idata는 반복요인의 데이터 프레임, idesign은 반복요인에 대한 식.

R Program (계속)

구형성 검정

```
summary(반복측정1.2, multivariate=T)$sphericity.tests
```

다변량 분석

```
summary(반복측정1.2, multivariate=T)$multivariate.tests
```

일변량 분석

```
summary(반복측정1.2, multivariate=T)$univariate.tests
```

```
summary(반복측정1.2, multivariate=T)$pval.adjustment
```

선형성 검토

```
qqnorm(residuals(반복측정1.1), main='Residuals')
```

```
qqline(residuals(반복측정1.1), lty=2)
```

○○○ 고혈압 자료의 구형성 검정 결과 ○○○

> # 구형성 검정

> summary(반복측정1.2, multivariate=T)\$sphericity.tests

	Test statistic	p-value
시간	0.74474	9.2926e-05
혈압강하제:시간	0.74474	9.2926e-05

시간, 혈압강하제와 시간의 교호작용 모두
구형성 가정을 만족하지 않음
-> 다변량 분석 방법이 적절함

고혈압 자료의 다변량 분석 결과

```
> summary(반복측정1.2, multivariate=T)$multivariate.tests
```

Multivariate Tests: 혈압강하제

	Df	test stat	approx F	num Df	den Df	Pr(>F)
Pillai	1	0.0224797	1.471787	1	64	0.22952
Wilks	1	0.9775203	1.471787	1	64	0.22952
Hotelling-Lawley	1	0.0229967	1.471787	1	64	0.22952
Roy	1	0.0229967	1.471787	1	64	0.22952

Multivariate Tests: 시간

	Df	test stat	approx F	num Df	den Df	Pr(>F)
Pillai	1	0.6795657	66.80409	2	63	2.6964e-16 ***
Wilks	1	0.3204343	66.80409	2	63	2.6964e-16 ***
Hotelling-Lawley	1	2.1207647	66.80409	2	63	2.6964e-16 ***
Roy	1	2.1207647	66.80409	2	63	2.6964e-16 ***

시간의 효과가 유의함

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Multivariate Tests: 혈압강하제:시간

	Df	test stat	approx F	num Df	den Df	Pr(>F)
Pillai	1	0.0950061	3.306863	2	63	0.043086 *
Wilks	1	0.9049939	3.306863	2	63	0.043086 *
Hotelling-Lawley	1	0.1049798	3.306863	2	63	0.043086 *
Roy	1	0.1049798	3.306863	2	63	0.043086 *

혈압강하제와 시간의
교호작용 효과가 유의함

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

○○○ 고혈압 자료의 일변량 분석 결과 ○○○

> # 일변량 검정

> summary(반복측정1.2, multivariate=T)\$univariate.tests

	Sum Sq	num Df	Error SS	den Df	F value	Pr(>F)
(Intercept)	1781063	1	6926.9	64	16455.8198	< 2e-16 ***
혈압강하제	159	1	6926.9	64	1.4718	0.22952
시간	5976	2	3744.0	128	102.1458	< 2e-16 ***
혈압강하제:시간	251	2	3744.0	128	4.2947	0.01566 *

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

시간의 효과가 유의함

혈압강하제와 시간의
교호작용 효과가 유의함

> summary(반복측정1.2, multivariate=T)\$pval.adjustment

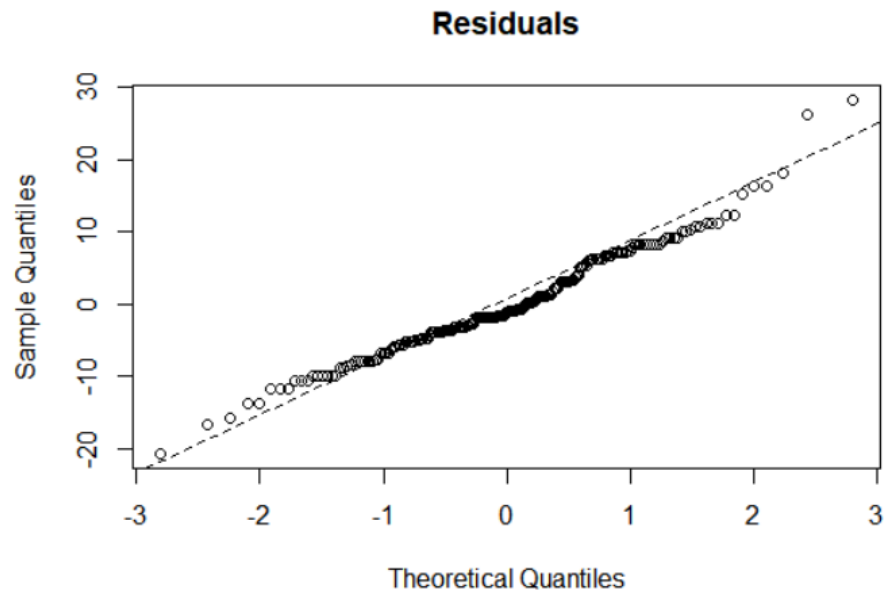
	GG eps	Pr(>F[GG])	HF eps	Pr(>F[HF])
시간	0.7966473	3.182882e-22	0.8137279	1.205013e-22
혈압강하제:시간	0.7966473	2.347864e-02	0.8137279	2.269227e-02

시간의 효과가 유의함

혈압강하제와 시간의
교호작용 효과가 유의함

선형성 검토 <그림 5.7>

```
> # 선형성 검토  
> qqnorm(residuals(반복측정1.1), main='Residuals')  
> qqline(residuals(반복측정1.1), lty=2)
```





반복요인이 둘 이상인 경우에 대한 반복측정 자료분석

반복요인이 둘 이상인 경우

- ▶ 반복요인이 둘 이상인 경우
 - ▶ 예 : 시간과 농도, 시간과 운동방법
- ▶ 반복요인이 2개인 자료
 - ➡ 반복요인이 하나인 반복측정모형을 그대로 확장

<예제 5.4>

- ◆ 남녀 심부전증 환자 각각 12명을 대상으로 한 임상시험에서 새로 나온 약에 대한 효과를 알고자 약의 농도를 1주일 간격으로 2mg, 3mg, 4mg 세 번 투여한 후 3시간마다 여덟 번을 반복하여 심전도를 측정하였다. 이 실험에서 얻은 자료는 <표 5.9>와 같다. 따라서 이 자료는 처리가 성별, 반복요인이 약의 농도와 시간으로 2개인 반복측정 설계가 된다.

<예제 5.4>

◆ 반복측정 자료분석에 대한 목표

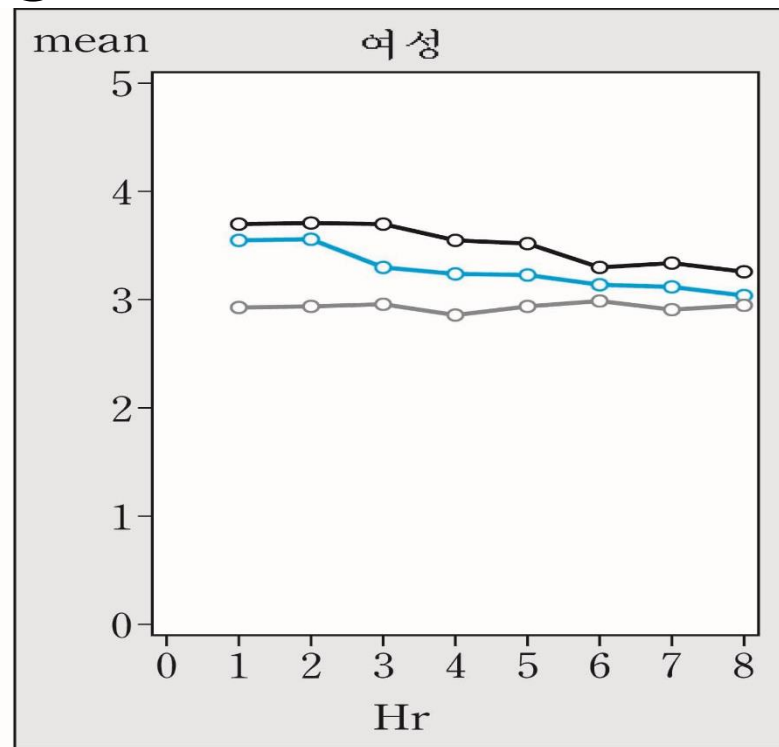
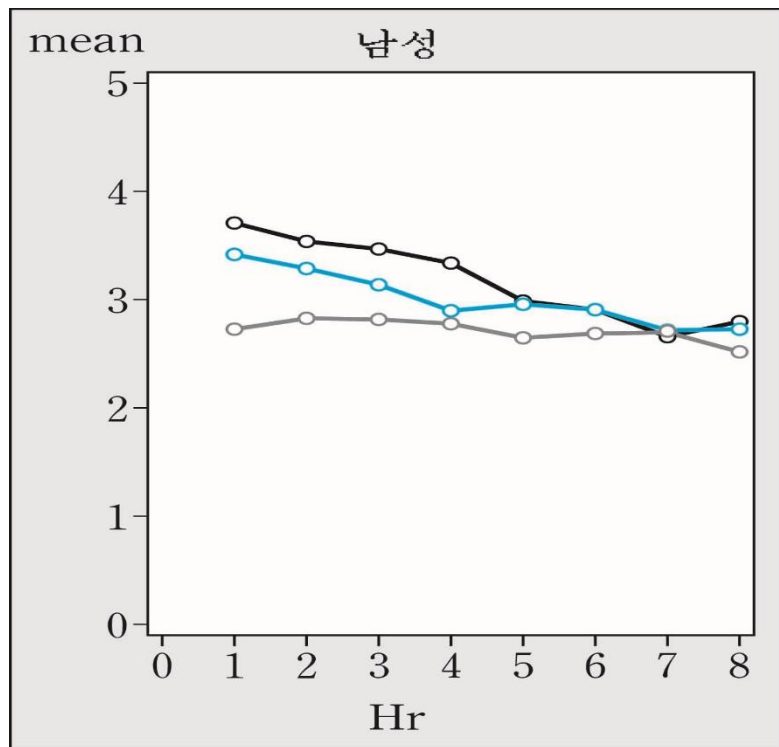
1. 각 반복요인에 따라 심전도의 차이가 있나?
2. 성별에 따라 심전도의 차이가 있나?
3. 성별과 반복요인들 간에 교호작용이 있나?

◆ 예제에서의 처리 : 성별

◆ 예제에서의 반복요인 : 약의 농도와 시간

시간에 따른 평균 심전도

<그림 5.8>



trt — 2mg — 3mg — 4mg

R Program

```
library(car)
options(contrasts=c("contr.sum", "contr.poly"))
심부전증환자자료=read.table("f:WWWWORKWW심부전증환자자료.txt",header=T)

# 반복측정 자료분석을 위한 데이터 변형
농도=rep(ordered(c("2mg", "3mg", "4mg")),each=8)
시간=rep(ordered(c(1:8)),3)
데이터2=data.frame(농도, 시간)
반복측정2.1 <- lm(as.matrix(심부전증환자자료[,-1]) ~ 성별, data= 심부전증환자자료)
반복측정2.2 <- Anova(반복측정2.1, idata=데이터2, idesign=~농도*시간, type='3')
```

- rep 함수** : 주어진 vector의 원소를 복제하여 새로운 vector를 만드는 함수. each = "x" 옵션을 사용한 경우, 각 원소를 x번 복제. each 옵션을 사용하지 않고 복제 횟수 x만 지정해 준 경우, 전체 vector를 x번 반복

R Program (계속)

구형성 검정

```
summary(반복측정2.2, multivariate=T)$sphericity.tests
```

다변량 검정

```
summary(반복측정2.2, multivariate=T)$multivariate.tests
```

일변량 검정

```
summary(반복측정2.2, multivariate=T)$univariate.tests  
summary(반복측정2.2, multivariate=T)$pval.adjustment
```

선형성 검토

```
qqnorm(residuals(반복측정2.1), main='Residuals')  
qqline(residuals(반복측정2.1), lty=2)
```

심부전증 환자 자료의 구형성 검정 결과

```
> summary(반복측정2.2, multivariate=T)$sphericity.tests
```

	Test statistic	p-value
농도	0.81909	0.12303
성별 : 농도	0.81909	0.12303
시간	0.00560	0.00000
성별 : 시간	0.00560	0.00000
농도 : 시간	0.00001	0.00000
성별 : 농도 : 시간	0.00001	0.00000



농도와 성별과 농도의 교호작용은 일변량 분석을,
나머지는 다변량 분석을 하는 것이 적절함.

심부전증 환자 자료의 다변량 분석 결과

```
> summary(반복측정2.2, multivariate=T)$multivariate.tests
```

Multivariate Tests: 성별

	Df	test stat	approx F	num Df	den Df	Pr(>F)
Pillai	1	0.0477633	1.103498	1	22	0.3049
Wilks	1	0.9522367	1.103498	1	22	0.3049
Hotelling-Lawley	1	0.0501590	1.103498	1	22	0.3049
Roy	1	0.0501590	1.103498	1	22	0.3049

Multivariate Tests: 농도

	Df	test stat	approx F	num Df	den Df	Pr(>F)
Pillai	1	0.7128294	26.06363	2	21	2.044e-06 ***
Wilks	1	0.2871706	26.06363	2	21	2.044e-06 ***
Hotelling-Lawley	1	2.4822501	26.06363	2	21	2.044e-06 ***
Roy	1	2.4822501	26.06363	2	21	2.044e-06 ***

Multivariate Tests: 성별:농도

	Df	test stat	approx F	num Df	den Df	Pr(>F)
Pillai	1	0.0196337	0.2102828	2	21	0.81204
Wilks	1	0.9803663	0.2102828	2	21	0.81204
Hotelling-Lawley	1	0.0200269	0.2102828	2	21	0.81204
Roy	1	0.0200269	0.2102828	2	21	0.81204

심부전증 환자 자료의 다변량 분석 결과 (계속)

Multivariate Tests: 시간

	Df	test stat	approx F	num Df	den Df	Pr(>F)
Pillai	1	0.762575	7.341373	7	16	0.0004966 ***
Wilks	1	0.237425	7.341373	7	16	0.0004966 ***
Hotelling-Lawley	1	3.211851	7.341373	7	16	0.0004966 ***
Roy	1	3.211851	7.341373	7	16	0.0004966 ***

Multivariate Tests: 성별:시간

	Df	test stat	approx F	num	Df	den	Df	Pr(>F)
Pillai	1	0.4266153	1.700639		7		16	0.1792
Wilks	1	0.5733847	1.700639		7		16	0.1792
Hotelling-Lawley	1	0.7440297	1.700639		7		16	0.1792
Rov	1	0.7440297	1.700639		7		16	0.1792

농도, 시간,
농도와 시간 교호작용
유의

Multivariate Tests: 농도:시간

	Df	test	stat	approx F	num Df	den Df	Pr(>F)
Pillai	1	0.860196	3.955416	14	9	0.021953	*
Wilks	1	0.139804	3.955416	14	9	0.021953	*
Hotelling-Lawley	1	6.152870	3.955416	14	9	0.021953	*
Roy	1	6.152870	3.955416	14	9	0.021953	*

Multivariate Tests: 성별:농도:시간

	Df	test stat	approx F	num Df	den Df	Pr(>F)
Pillai	1	0.6831168	1.38583	14	9	0.31677
Wilks	1	0.3168832	1.38583	14	9	0.31677
Hotelling-Lawley	1	2.1557364	1.38583	14	9	0.31677
Roy	1	2.1557364	1.38583	14	9	0.31677

심부전증 환자 자료의 일변량 분석 결과

```
> summary(반복측정2.2, multivariate=T)$univariate.tests
```

	Sum Sq	num Df	Error SS	den Df	F value	Pr(>F)
(Intercept)	5481.9	1	211.391	22	570.5112	< 2.2e-16 ***
성별	10.6	1	211.391	22	1.1035	0.304903
농도	26.2	2	24.166	44	23.8959	9.425e-08 ***
성별:농도	0.3	2	24.166	44	0.2480	0.781482
시간	17.0	7	10.614	154	35.1649	< 2.2e-16 ***
성별:시간	1.8	7	10.614	154	3.7510	0.000882 ***
농도:시간	6.3	14	18.305	308	7.5274	1.362e-13 ***
성별:농도:시간	1.2	14	18.305	308	1.4054	0.148783

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```
> summary(반복측정2.2, multivariate=T)$pval.adjustment
```

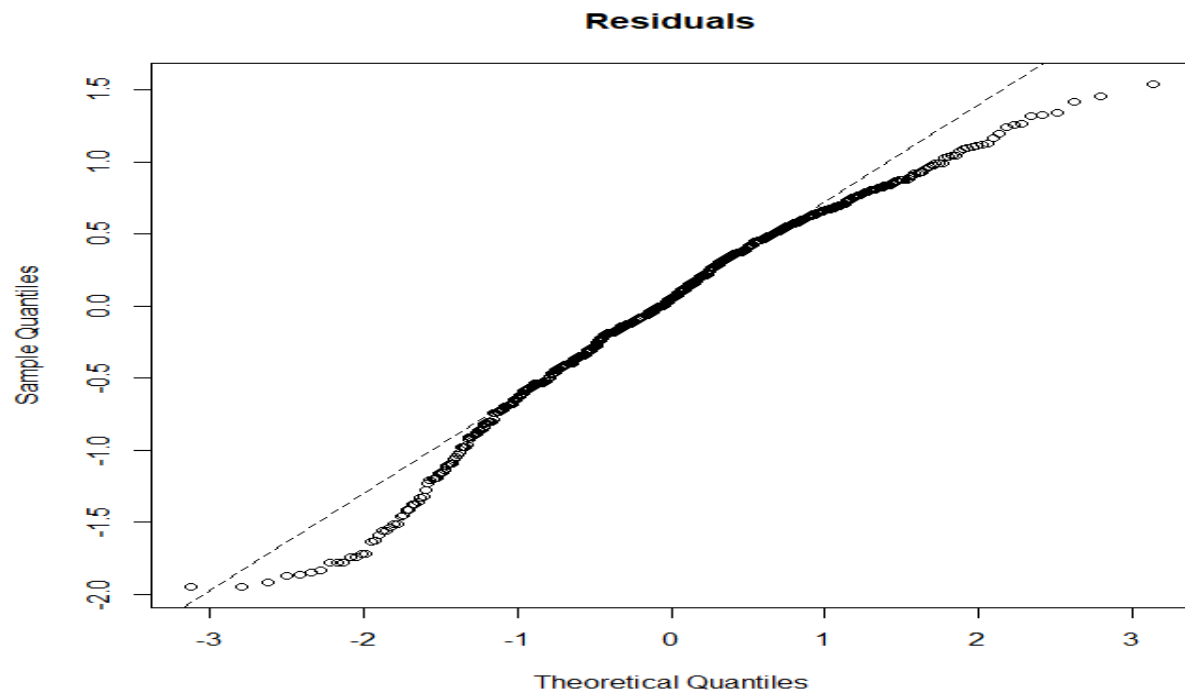
	GG eps	Pr(>F[GG])	HF eps	Pr(>F[HF])
농도	0.8468081	7.037570e-07	0.9098905	3.072376e-07
성별:농도	0.8468081	7.449848e-01	0.9098905	7.609290e-01
시간	0.3532950	1.131648e-11	0.4015003	5.714060e-13
성별:시간	0.3532950	2.205760e-02	0.4015003	1.721004e-02
농도:시간	0.4137582	8.862847e-07	0.5783524	1.046793e-08
성별:농도:시간	0.4137582	2.193669e-01	0.5783524	1.961409e-01

일변량 분석의 결과 농도, 시간, 성별과 시간의 교호작용,
농도와 시간의 교호작용이 유의함

심부전증 환자 자료의 일변량 분석과 다변량 분석 결과 비교 <표 5.10>

구형성 조건	요인	다변량 분석	일변량 분석
농도(만족)	농도	유의	유의
	농도*성별	유의하지 않음	유의하지 않음
시간(불만족)	시간	유의	유의
	시간*성별	유의하지 않음	유의
농도*시간(불만족)	농도*시간	유의	유의
	농도*시간*성별	유의하지 않음	유의하지 않음
	성별	유의하지 않음	유의하지 않음

심부전증 환자 자료의 선형성 검토 <그림 5.9>



결측치 처리

- ▶ 반복측정 자료분석은 여러 시간 혹은 공간에 걸쳐 반복적으로 측정되는 실험에 유용한 통계분석 방법임.
- ▶ 자료에 결측치가 많이 존재하는 경우에는 분석 방법 적용에 어려움이 있음.
- ▶ 결측치를 대체하는 방법
 - ▶ 방법 1 : 결측치가 있는 시간 이전 시간의 값으로 결측치를 대체.
 - ▶ 방법 2 : 인접한 값들의 평균을 사용.

핵심정리

1. 반복측정 자료분석 : 한 개체에 대해 반복측정 한 자료에 대한 분석.
2. 오차항의 구형성 : 반복측정 자료에 대해 일변량 분석을 하기 위해 만족해야 하는 가정 중 하나로 반복요인의 오차항의 동일한 공분산.
3. 연구가설

* 반복요인이 하나인 경우

책 pg.227 요약 부분 설명이
이와 같아야 합니다.

$$\begin{aligned} H_{01} : \alpha_1 = \alpha_2 = \cdots = \alpha_I = 0 & \quad (\text{처리 간의 반응측정치의 차이가 없다.}) \\ H_{02} : \gamma_1 = \gamma_2 = \cdots = \gamma_K = 0 & \quad (\text{시간 간의 반응측정치의 차이가 없다.}) \\ H_{03} : \delta_{11} = \delta_{12} = \cdots = \delta_{IK} = 0 & \quad (\text{처리와 시간 간의 교호작용이 존재하지 않는다.}) \end{aligned}$$



제6강

수고하셨습니다!

