

Part 2. R 통계분석 (데이터 분석 전문가 양성과정)

11

반복측정과 공분산 분석

경북대학교 배준현 교수
(joonion@knu.ac.kr)



11. 반복측정과 공분산 분석

- 분산분석: analysis of variance (*ANOVA*)
 - 일원 분산분석: *one-way* ANOVA
 - 이원 분산분석: *two-way* ANOVA
 - 반복측정 분산분석: *repeated-measures* ANOVA
 - 공분산 분석: analysis of covariance (*ANCOVA*)
 - 다변량 분산분석: *multivariate* analysis of variance (*MANOVA*)



11. 반복측정과 공분산 분석

- **반복측정 분산분석:** *repeated measures* ANOVA
 - 하나의 실험 대상에 대해 두 번 이상 반복측정한 관측값 간의 차이 검정
 - 반복측정 이원 분산분석:
 - 집단 간 요인과 집단 내 요인이 각각 한 개씩일 때
 - 집단 간 요인과 집단 내 요인이 모두 집단변수로서의 역할 수행
 - 집단 내 요인의 각 범주별로 관측값을 반복측정



11. 반복측정과 공분산 분석

- 표준패키지의 CO₂ 데이터셋: 식물이 저온의 생장환경에서 견디는 정도에 대한 실험 결과

```
> str(CO2)
```

Classes 'nfnGroupedData', 'nfGroupedData', 'groupedData' and 'data.frame':

84 obs. of 5 variables:

```
$ Plant      : Ord.factor w/ 12 levels "Qn1"<"Qn2"<"Qn3"<...: 1 1 1 1 1 1 1 2 2 2 ...
$ Type       : Factor w/ 2 levels "Quebec", "Mississippi": 1 1 1 1 1 1 1 1 1 1 ...
$ Treatment: Factor w/ 2 levels "nonchilled", "chilled": 1 1 1 1 1 1 1 1 1 1 ...
$ conc       : num  95 175 250 350 500 675 1000 95 175 250 ...
$ uptake     : num  16 30.4 34.8 37.2 35.3 39.2 39.7 13.6 27.3 37.1 ...
- attr(*, "formula")=Class 'formula'  language uptake ~ conc | Plant
.. ..- attr(*, ".Environment")=<environment: R_EmptyEnv>
- attr(*, "outer")=Class 'formula'  language ~Treatment * Type
.. ..- attr(*, ".Environment")=<environment: R_EmptyEnv>
- attr(*, "labels")=List of 2
..$ x: chr "Ambient carbon dioxide concentration"
..$ y: chr "CO2 uptake rate"
- attr(*, "units")=List of 2
..$ x: chr "(uL/L)"
..$ y: chr "(umol/m^2 s)"
```



11. 반복측정과 공분산 분석

```
> head(CO2, 3); tail(CO2, 3)
```

```
Grouped Data: uptake ~ conc | Plant
```

	Plant	Type	Treatment	conc	uptake
1	Qn1	Quebec	nonchilled	95	16.0
2	Qn1	Quebec	nonchilled	175	30.4
3	Qn1	Quebec	nonchilled	250	34.8

```
Grouped Data: uptake ~ conc | Plant
```

	Plant	Type	Treatment	conc	uptake
82	Mc3	Mississippi	chilled	500	17.9
83	Mc3	Mississippi	chilled	675	18.9
84	Mc3	Mississippi	chilled	1000	19.9



11. 반복측정과 공분산 분석

- 반복측정 분산분석을 위한 포물러 형식:
 - 반복측정 **일원** 분산분석: $y \sim W + \text{Error}(\text{Subject}/W)$
 - 반복측정 **이원** 분산분석: $y \sim B * W + \text{Error}(\text{Subject}/W)$
 - W: 집단 내 요인, B: 집단 간 요인
 - Subject: 각 측정 대상에 대한 식별자 변수



11. 반복측정과 공분산 분석

```
> CO2sub <- subset(CO2, Treatment=="chilled")
> CO2sub$conc <- factor(CO2sub$conc)
> CO2sub.aov <- aov(uptake ~ Type * conc + Error(Plant/conc), data=CO2sub)
> summary(CO2sub.aov)
```

Error: Plant

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Type	1	2667.2	2667.2	60.41	0.00148 **
Residuals	4	176.6	44.1		

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

Error: Plant:conc

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
conc	6	1472.4	245.40	52.52	1.26e-12 ***
Type:conc	6	428.8	71.47	15.30	3.75e-07 ***
Residuals	24	112.1	4.67		

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



11. 반복측정과 공분산 분석

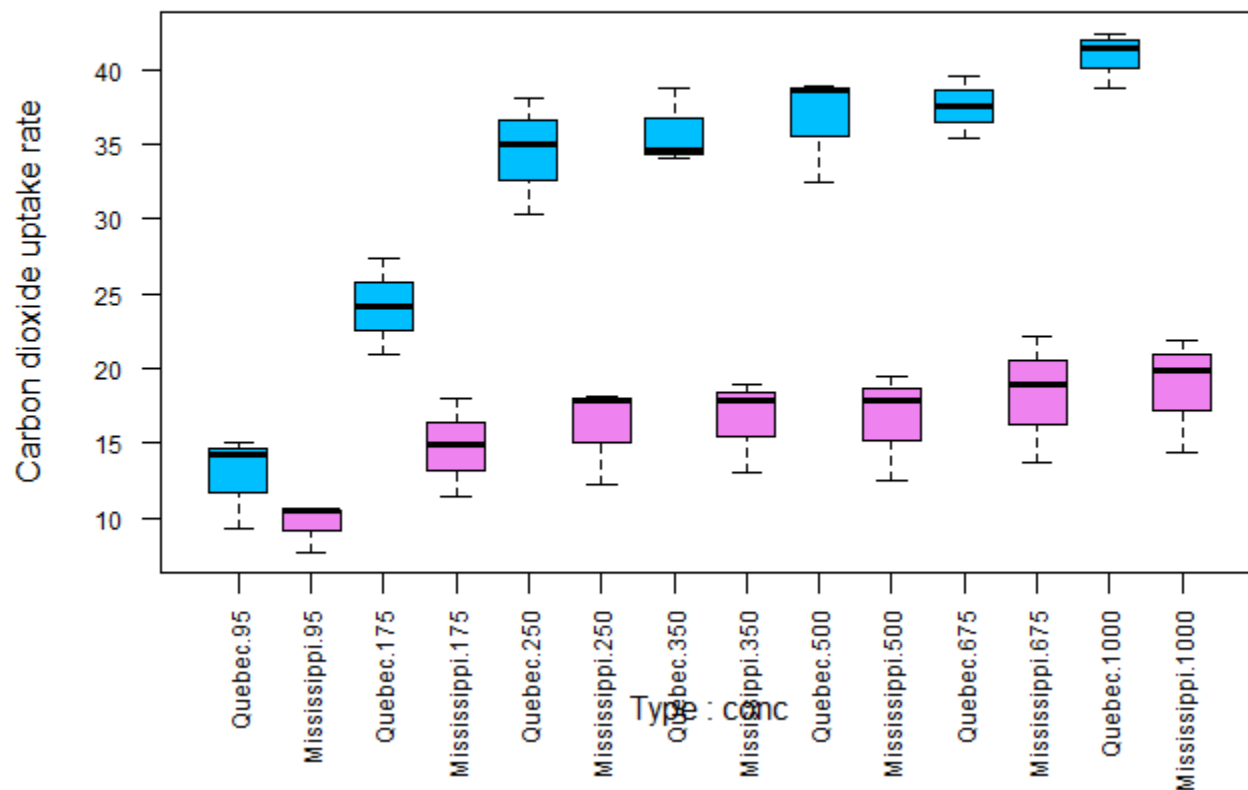
- 상자도표를 통해 주 효과와 상호작용 효과를 확인

```
par(mar=c(6,4,4,2))
boxplot(uptake ~ Type * conc, data=C02sub,
        col=c("deepskyblue", "violet"), las=2, cex.axis=0.75,
        ylab="Carbon dioxide uptake rate",
        main="Effects of Plant Type and CO2 on Carbon Dioxide Uptake")
legend("topleft", inset=0.02,
       legend=c("Quebec", "Mississippi"), fill=c("deepskyblue", "violet"))
```




11. 반복측정과 공분산 분석

Effects of Plant Type and CO₂ on Carbon Dioxide Uptake

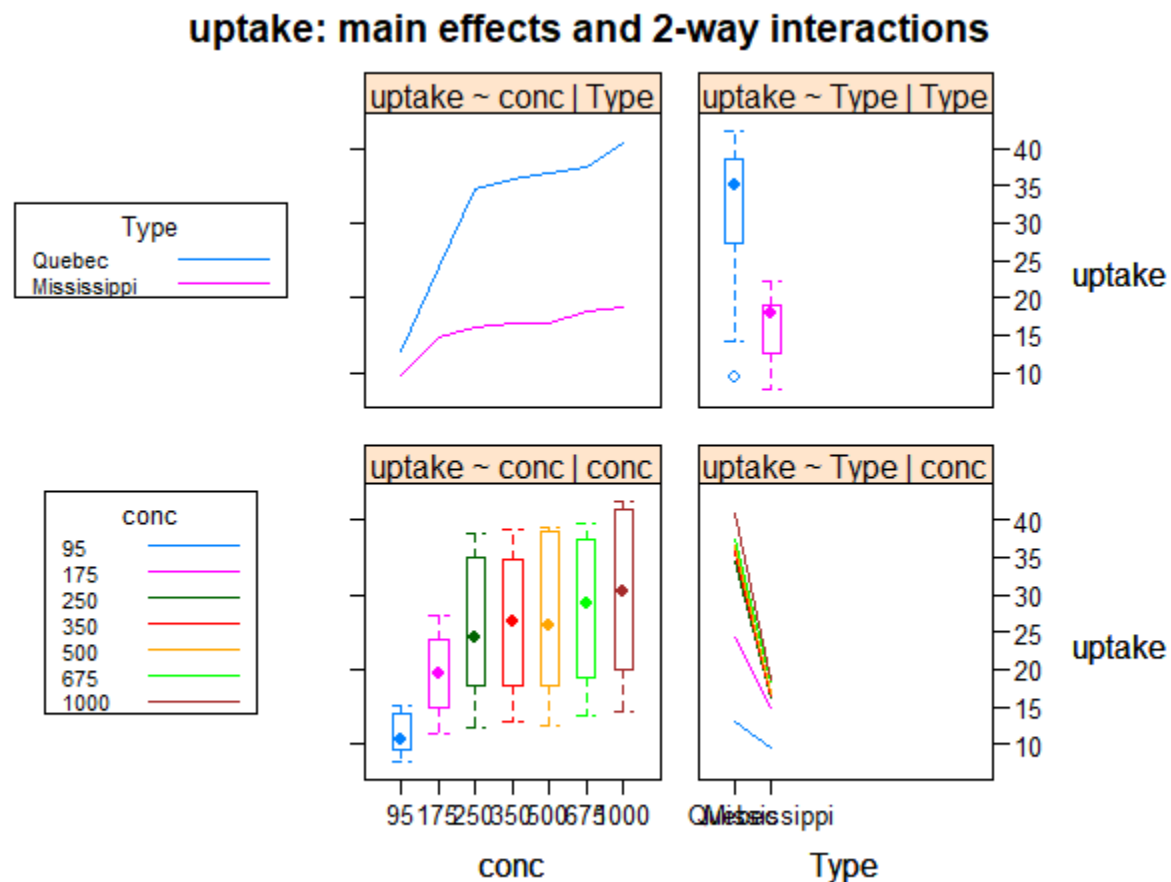




11. 반복측정과 공분산 분석

- HH 패키지의 `interaction2wt()` 함수를 이용하여 주 효과와 상호작용 효과를 확인

```
library(HH)
interaction2wt(uptake ~ conc * Type, data=C02sub)
```





11. 반복측정과 공분산 분석

- 공분산 분석: *ANCOVA* (analysis of covariance)
 - 분산분석에 공변량을 추가하여 분산분석 모델을 확장



11. 반복측정과 공분산 분석

- faraway 패키지의 **sexab 데이터셋**: 아동기의 성폭력 경험이 성인의 정신건강에 미치는 영향 연구

```
> library(faraway)
```

```
> str(sexab)
```

```
'data.frame': 76 obs. of 3 variables:
```

```
$ cpa : num 2.048 0.839 -0.241 -1.115 2.015 ...
```

```
$ ptsd: num 9.71 6.17 15.16 11.31 9.95 ...
```

```
$ csa : Factor w/ 2 levels "Abused","NotAbused": 1 1 1 1 1 1 1 1 1 1 ...
```



11. 반복측정과 공분산 분석

- 집단별(성폭력 경험 유무별) 요약통계량 확인

```
> tapply(sexab$ptsd, sexab$csa, mean)
```

```
Abused NotAbused  
11.941093  4.695874
```

```
> tapply(sexab$ptsd, sexab$csa, sd)
```

```
Abused NotAbused  
3.440152  3.519743
```

```
> tapply(sexab$ptsd, sexab$csa, length)
```

```
Abused NotAbused  
45      31
```



11. 반복측정과 공분산 분석

- 일원 공분산 분석을 통해 아동기의 성폭력 경험과 외상 후 스트레스 장애 간의 관계를 통계적 검정

```
> sexab.aov <- aov(ptsd ~ cpa + csa, data=sexab)
```

```
> summary(sexab.aov)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
cpa	1	449.8	449.8	41.98	9.46e-09 ***
csa	1	624.0	624.0	58.25	6.91e-11 ***
Residuals	73	782.1	10.7		

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



11. 반복측정과 공분산 분석

- effects 패키지의 effects() 함수: 공변량의 영향을 제거한 후 외상 후 스트레스 장애의 집단 평균 계산

```
> library(effects)
> effect("csa", sexab.aov)
```

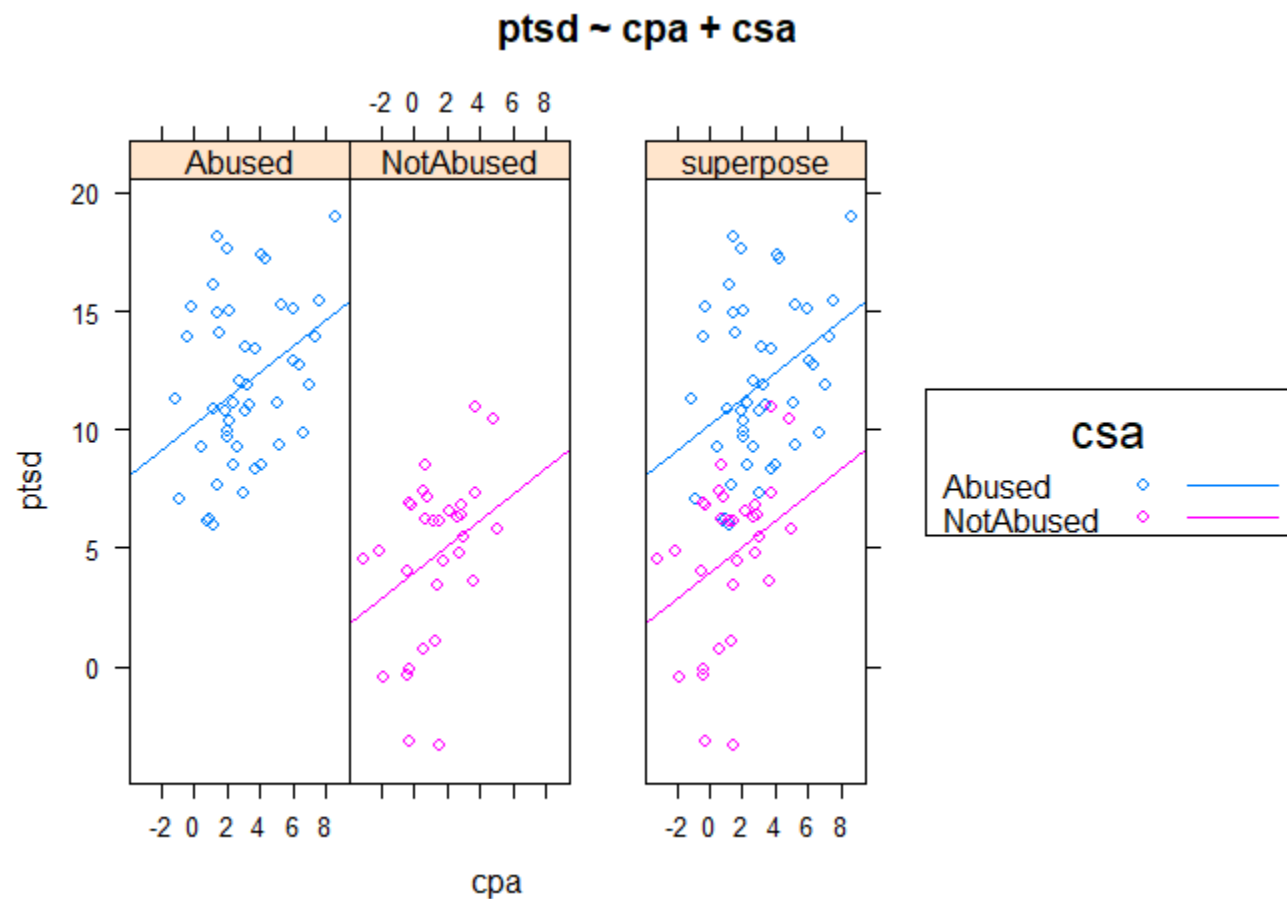
```
csa effect
csa
  Abused NotAbused
11.544429  5.271677
```



11. 반복측정과 공분산 분석

- HH 패키지의 `ancova()` 함수: 종속변수, 공변량, 독립변수 간의 관계 그래프

```
> library(HH)  
> ancova(ptsd ~ cpa + csa, data = sexab)
```





11. 반복측정과 공분산 분석

- 다변량 분산분석: MANOVA (multivariate ANOVA)
 - 두 개 이상의 종속변수가 있을 경우 다변량 분산분석을 이용



11. 반복측정과 공분산 분석

- heplots 패키지의 **Skulls 데이터셋**: 이집트 지역에서 발굴된 인간의 두개골 크기 측정

```
> library(heplots)
```

```
> str(Skulls)
```

```
'data.frame': 150 obs. of 5 variables:
```

```
$ epoch: Ord.factor w/ 5 levels "c4000BC"<"c3300BC"<...: 1 1 1 1 1 1 1 1 1 1 ...
```

```
$ mb : num 131 125 131 119 136 138 139 125 131 134 ...
```

```
$ bh : num 138 131 132 132 143 137 130 136 134 134 ...
```

```
$ bl : num 89 92 99 96 100 89 108 93 102 99 ...
```

```
$ nh : num 49 48 50 44 54 56 48 48 51 51 ...
```



11. 반복측정과 공분산 분석

```
> library(dplyr)
> sample_n(Skulls, 10)
```

	epoch	mb	bh	b1	nh
3	c4000BC	131	132	99	50
104	c200BC	133	136	95	52
11	c4000BC	129	138	95	50
149	cAD150	147	129	87	48
87	c1850BC	136	137	101	54
53	c3300BC	135	136	97	52
74	c1850BC	137	133	90	49
59	c3300BC	135	132	98	54
134	cAD150	137	134	93	53
139	cAD150	133	125	92	50



11. 반복측정과 공분산 분석

- 다변량 분산분석을 통해 두개골의 측정값이 시대에 따라 다른지 검정

```
> y <- with(Skulls, cbind(mb, bh, bl, nh))  
> with(Skulls, aggregate(y, by=list(epoch), FUN=mean))
```

	Group.1	mb	bh	bl	nh
1	c4000BC	131.3667	133.6000	99.16667	50.53333
2	c3300BC	132.3667	132.7000	99.06667	50.23333
3	c1850BC	134.4667	133.8000	96.03333	50.56667
4	c200BC	135.5000	132.3000	94.53333	51.96667
5	cAD150	136.1667	130.3333	93.50000	51.36667



11. 반복측정과 공분산 분석

- manova() 함수: 종속변수값(두개골 크기)의 집단 간 차이(시대별 차이)에 대한 다변량 분산분석 수행

```
> Skulls.manova <- with(Skulls, manova(y ~ epoch))
```

```
> summary(Skulls.manova)
```

	Df	Pillai	approx	F	num	Df	den	Df	Pr(>F)
epoch	4	0.35331	3.512		16		580		4.675e-06 ***

Residuals 145

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



11. 반복측정과 공분산 분석

- summary.aov() 함수: 일변량 일원분산분석으로 어느 두개골 측정값에 차이가 존재하는지 확인

```
> summary.aov(Skulls.manova)
```

Response **mb** :

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
epoch	4	502.83	125.707	5.9546	0.0001826 ***
Residuals	145	3061.07	21.111		

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

Response **bh** :

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
epoch	4	229.9	57.477	2.4474	0.04897 *
Residuals	145	3405.3	23.485		

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

Response **bl** :

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
epoch	4	803.3	200.823	8.3057	4.636e-06 ***
Residuals	145	3506.0	24.179		

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

Response **nh** :

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
epoch	4	61.2	15.300	1.507	0.2032
Residuals	145	1472.1	10.153		

Any Questions?

