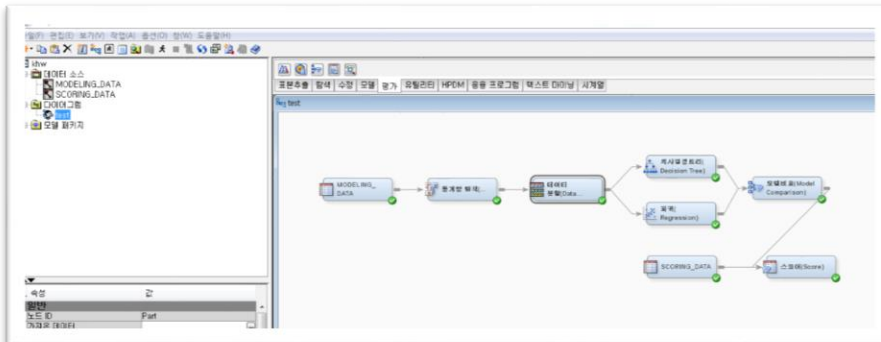


실험계획법 실습 - 14 주차

기말고사 대비 및 E-miner 실습

```
proc glm data=bibd;  
class treat block;  
model y=treat block;  
means treat block/tukey;  
run;
```



로그아웃 회원정보수정 SAS 코리아 바로가기 SAS t f

[대회 소개](#) [공모요강 및 시상내역](#) [소프트웨어 및 교육](#) [공지사항](#) [Q&A](#) [FAQ](#)

Big Data 시대의 Data Scientist를 향한 도전

건강보험심사평가원과 함께하는
제12회 SAS 마이닝 챔피언십 [등록하기](#)

Big Data 시대의 Data Scientist를 꿈꾸는 여러분!!
제 12회 SAS 마이닝 챔피언십에 도전하세요!

Support Your Dream

“Big Data 시대의 Data Scientist를 향한 도전”

SAS 코리아가 Big Data 시대를 이끌어갈 Data Scientist 양성을 위해
“제 12회 SAS 마이닝 챔피언십”을 개최합니다.

최근들어 세계 선진 기업들은 소셜미디어와 스마트기기의 확산으로 정형데이터 뿐만 아니라
비정형데이터까지 폭증하는 데이터 시대, 즉 Big Data의 시대를 대비하기 위해
Data Miner를 넘어 Data Scientist를 찾고 있고 양성해야 한다는 목소리가 높아지고 있습니다.

이에 세계적인 비즈니스 분석 소프트웨어 부문 리더인 SAS에서는 12번째 SAS 마이닝 챔피언십을 개최합니다.
10년 넘게 우수한 분석가들을 발굴해온 SAS 마이닝 챔피언십은 이제 단순 Miner의 발굴을 넘어서
Data Scientist 양성을 위해 앞장서고 있습니다.

2014 제 12회 SAS 마이닝 챔피언십에서 데이터 분석에 대한 진지한 고민,
차별화된 아이디어로 Data Scientist 로 나가길 대학생 및 대학원생 여러분의 많은 참여를 기대합니다.



주관: 한국제스소프트웨어(유) 후원: 건강보험심사평가원 지원: 한국건강관리협회

The Randomized Complete Block Design

● Data

블록 1	자동차			
	1	2	3	4
	C(12)	A(14)	D(10)	A(13)
	A(17)	A(13)	C(11)	D(9)
	D(13)	B(14)	B(14)	B(8)
	D(11)	C(12)	B(13)	C(9)

블록 2	자동차			
	1	2	3	4
	B(14)	D(11)	A(13)	C(9)
	C(12)	C(12)	B(13)	D(9)
	A(17)	B(14)	D(11)	B(8)
	D(13)	A(14)	C(10)	A(13)

The Randomized Complete Block Design

- 랜덤화완비블록설계(randomized complete block desing, RCBD)

- 모형: $y_{ij} = \mu + \tau_i + \beta_j + \varepsilon_{ij}, \varepsilon_{ij} \sim NID(0, \sigma^2)$

- τ_i : 처리 i 에 대한 효과. ($i = 1, 2, 3, 4$)

- β_j : 블록 j 에 대한 효과. ($j = 1, 2, 3, 4$)

- $\sum_{i=1}^4 \tau_i = 0$ & $\sum_{j=1}^4 \beta_j = 0$.

여기서는 차(블록)에 기인한 변동이 분리하여, 블록효과에 대한 검정이 가능하지만, 주 목적은 브랜드 차이의 검정에 있다.

따라서 가설의 경우는 $H_0: \tau_1 = \tau_2 = \tau_3 = \tau_4 = 0$ vs $H_1: \text{not } H_0$

The Latin Square Design

● Data

브랜드	자동차			
	1	2	3	4
	B(14)	D(11)	A(13)	C(9)
	C(12)	C(12)	B(13)	D(9)
	A(17)	B(14)	D(11)	B(8)
	D(13)	A(14)	C(10)	A(13)

위치	자동차			
	1	2	3	4
1	C(12)	D(11)	A(13)	B(8)
2	B(14)	C(12)	D(11)	A(13)
3	A(17)	B(14)	C(10)	D(9)
4	D(13)	A(14)	B(13)	C(9)

The Latin Square Design

- Latin Square Design

- 모형: $y_{ijk} = \mu + \tau_i + \beta_j + \gamma_k + \varepsilon_{ijk}, \varepsilon_{ijk} \sim NID(0, \sigma^2)$
- τ_i : 처리 i 에 대한 효과. ($i = 1, 2, 3, 4$)
- β_j : j 번째 블록(열)에 대한 효과. ($j = 1, 2, 3, 4$): 차에 대한 효과
- γ_k : k 번째 블록(행)에 대한 효과. ($k = 1, 2, 3, 4$): 위치에 대한 효과

Balanced Incomplete Block Designs

- 예시: 처리는 3개이지만 한 블록에 2개 처리만 가능한 경우

처리	블록		
	1	2	3
A	A	.	A
B	B	B	.
C	.	C	C

- 예시: 처리는 4개이지만 한 블록에 2개 처리만 가능한 경우

처리	블록					
	1	2	3	4	5	6
A	A	A	A	.	.	.
B	B	.	.	B	B	.
C	.	C	.	C	.	C
D	.	.	D	.	D	D

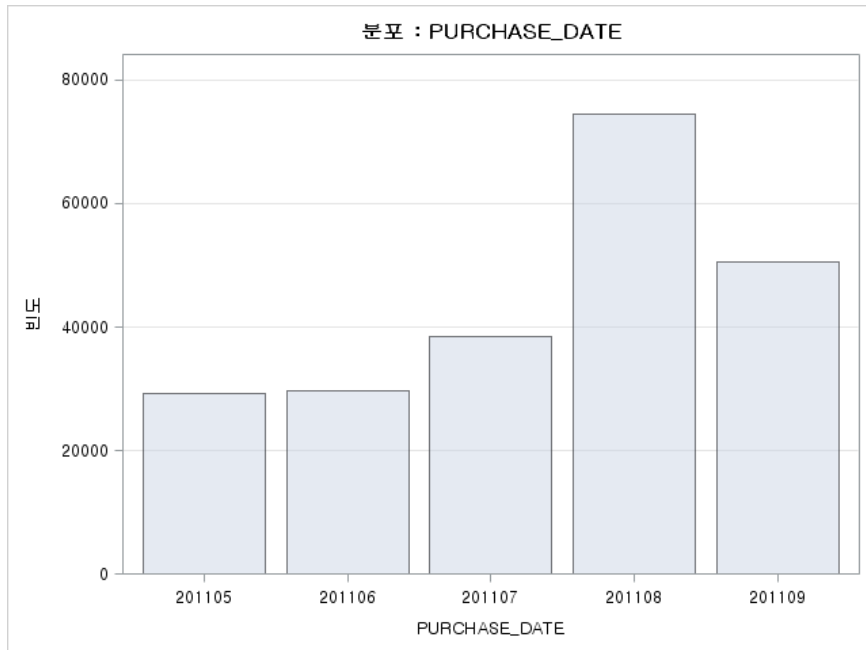
Balanced Incomplete Block Designs

- BIBD

- 모형: $y_{ij} = \mu + \tau_i + \beta_j + \varepsilon_{ij}, \varepsilon_{ij} \sim NID(0, \sigma^2)$
- τ_i : 처리 i 에 대한 효과. ($i = 1, 2, 3, \dots, a$)
- β_j : j 번째 블록(열)에 대한 효과. ($j = 1, 2, 3, \dots, b$)
- 랜덤화 블록 설계(RCBD)와 모형은 같지만 다른 점?
 - 1) 일부 y_{ij} 에 대한 값이 존재 하지 않음.
 - 2) $a \times b \neq$ 총관측치수
 - 3) $SST = SS_{block} + SS_{treatment} + SSE$ 의 차이!!

E-miner 실습

● 재 구매 고객 모델링



- 5월 ~ 9월 데이터를 가지고 10월에 재 구매 할 고객 10,000명을 예측

E-miner 실습

● 재 구매 고객 모델링 변수 설명

변수	형태	설명
CUST_ID	숫자	고객 ID
AGE	숫자	나이
JOB_CD	문자	직업
sex	문자	성별
COUNT_of_CUST_ID	숫자	구매횟수
SUM_of_PRICE	숫자	총 금액
SUM_of_BUY_5	숫자	5월 달에 구매한 횟수
SUM_of_BUY_6	숫자	6월 달에 구매한 횟수
SUM_of_BUY_7	숫자	7월 달에 구매한 횟수
SUM_of_BUY_8	숫자	8월 달에 구매한 횟수
SUM_of_BUY_A	숫자	A 구매한 횟수
SUM_of_BUY_B	숫자	B 구매한 횟수
SUM_of_BUY_C	숫자	C 구매한 횟수
Target	숫자	5~8월 달에 구매를 하고 9월에 구매했으면 타겟 1

E-miner 실습

● Modeling_Data 변수 설정

이름	역할	레벨	리포트	순서	제거	하한	상한
AGE	Input	Interval	아니요		아니요	.	.
COUNT_of_CUST_ID	Input	Interval	아니요		아니요	.	.
CUST_ID	ID	Nominal	아니요		아니요	.	.
JOB_CD	Input	Nominal	아니요		아니요	.	.
sex	Input	BINARY	아니요		아니요	.	.
SUM_of_BUY_5	Input	Interval	아니요		아니요	.	.
SUM_of_BUY_6	Input	Interval	아니요		아니요	.	.
SUM_of_BUY_7	Input	Interval	아니요		아니요	.	.
SUM_of_BUY_8	Input	Interval	아니요		아니요	.	.
SUM_of_BUY_A	Input	Interval	아니요		아니요	.	.
SUM_of_BUY_B	Input	Interval	아니요		아니요	.	.
SUM_of_BUY_C	Input	Interval	아니요		아니요	.	.
SUM_of_PRICE	Input	Interval	아니요		아니요	.	.
Target	Target	BINARY	아니요		아니요	.	.

E-miner 실습

● Scoring_Data 변수 설정(역할: Score)

이름	역할	레벨	리포트	순서	제거	하한	상한
AGE	Input	Interval	아니요		아니요	.	.
COUNT_of_CUST_ID	Input	Interval	아니요		아니요	.	.
CUST_ID	ID	Nominal	아니요		아니요	.	.
JOB_CD	Input	Nominal	아니요		아니요	.	.
sex	Input	BINARY	아니요		아니요	.	.
SUM_of_BUY_5	Input	Interval	아니요		아니요	.	.
SUM_of_BUY_6	Input	Interval	아니요		아니요	.	.
SUM_of_BUY_7	Input	Interval	아니요		아니요	.	.
SUM_of_BUY_8	Input	Interval	아니요		아니요	.	.
SUM_of_BUY_A	Input	Interval	아니요		아니요	.	.
SUM_of_BUY_B	Input	Interval	아니요		아니요	.	.
SUM_of_BUY_C	Input	Interval	아니요		아니요	.	.
SUM_of_PRICE	Input	Interval	아니요		아니요	.	.
Target	Rejected	Interval	아니요		아니요	.	.

E-miner 실습

● 노드 구성 및 실행

