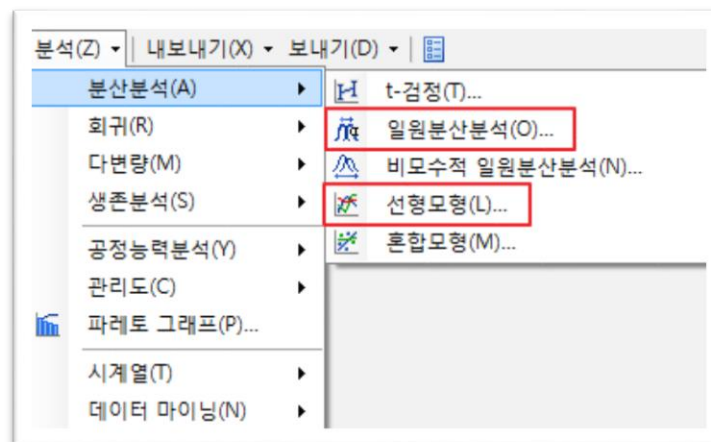
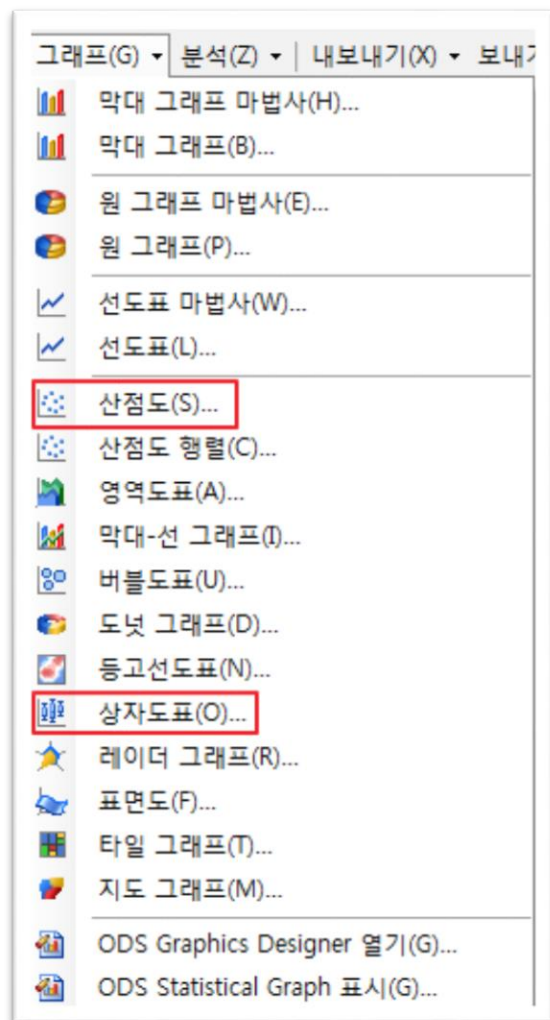


실험계획법 실습 - 5주차

Chapter3 Analysis of Variance(ANOVA)



```
libname inha "C:\Users\Wkhw\Desktop";

data ex1;
  set inha.ex1;

  /*가변수 작업*/

  if treat="A" then treat_A=1;
  else treat_A=0;

  if treat="B" then treat_B=1;
  else treat_B=0;

run;
```

The Analysis of Variance

- The basic single-factor ANOVA Model

$$y_{ij} = \mu + \tau_i + \varepsilon_{ij}, \begin{cases} i = 1, 2, \dots, a \\ j = 1, 2, \dots, n \end{cases}$$

- y_{ij} : i 번째 수준에서의 j 번째 관측한 값.
- μ : 전체 관측값의 평균(모평균).
- τ_i : i 번째 수준의 효과.
- ε_{ij} : i 번째 수준에서의 j 번째 관측 y_{ij} 에 수반되는 오차($\varepsilon_{ij} \sim NID(0, \sigma^2)$).

The Analysis of Variance

- Model Assumptions

- 1) Fixed Effect Model: $H_0: \tau_1 = \tau_2 = \dots = \tau_a = 0$

- $\tau_i: i=1, 2, 3, \dots, a$ are fixed parameters.
 - $\sum_{i=1}^a \tau_i = 0 \rightarrow \mu = \sum_{i=1}^a \mu_i / a$, where $\mu_i = \mu + \tau_i$.

- 2) Random Effect Model: $H_0: \sigma_\tau^2 = 0$ (책 3.9 The Random Effects Model참고)

- $\tau_i: i=1, 2, 3, \dots, a$ are random variables.
 - $\tau_i \sim NID(0, \sigma_\tau^2)$.
 - 책 Page 69 Fixed or Random Factor? 부분 참고
 - 평균을 비교하기 힘들기 때문에!!

The Analysis of Variance

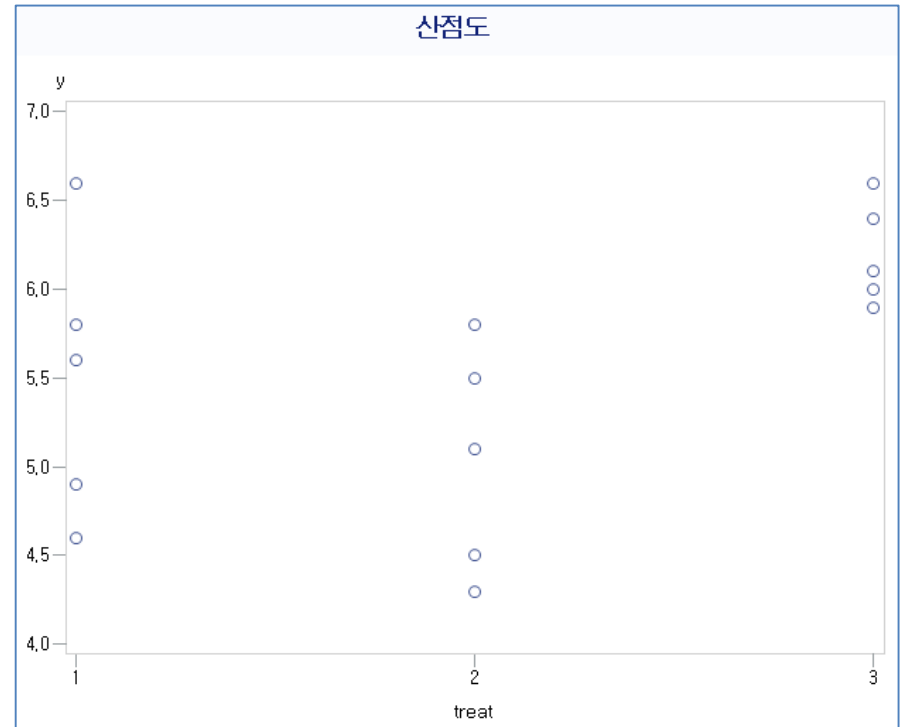
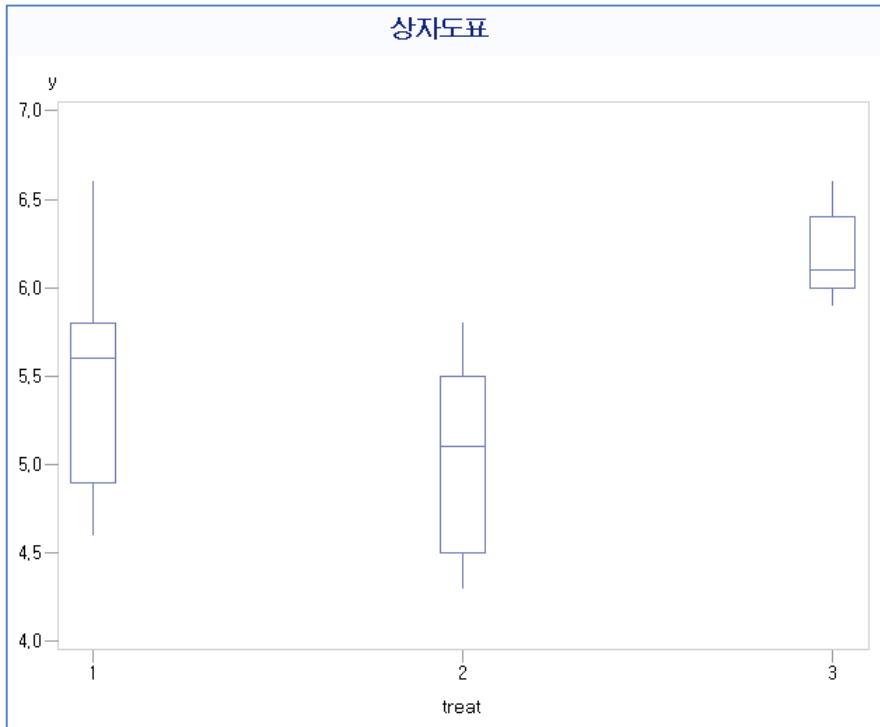
● EX1) 다이어트 약

다이어트 약			
OBS	A(1)	B(2)	C(3)
1	4.6	4.3	6.4
2	5.8	5.8	6.6
3	5.6	5.1	5.9
4	4.9	5.5	6.0
5	6.6	4.5	6.1

- 3가지 다이어트 약의 효능을 비교하고자 한다.
과 체중 남성 15명을 Random하게 선택하고,
3 집단에 5명씩 Random하게 배정하였다.
실험 전에 신체검사를 하여 집단 별로
과 체중의 평균을 비교한 결과 유의한 차이가
없었다. 일정 기간 약을 복용한 후, 체중 감소량(kg)을
측정하여 다음의 데이터를 얻었다.

The Analysis of Variance

- EX1을 가지고 약에 따른 box plot 및 Scatter diagram을 그려보세요.



The Analysis of Variance

- EX1에서 모형(Fixed간주) 및 가정을 쓰고, 가설을 작성하세요.

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

τ_i : the i th treatment effect (i 번째 수준에서의 효과) & $\sum_{i=1}^3 \tau_i = 0$

단, $i = 1, 2, 3$ $j = 1, 2, 3, 4, 5$

- 가설: $H_0: \tau_1 = \tau_2 = \tau_3 = 0$ vs $H_1: \text{not } H_0$

The Analysis of Variance

- EX1를 유의수준 0.05에서 ANOVA 분석을 하시오.

일원분산분석
결과
The ANOVA Procedure

Class Level Information		
Class	Levels	Values
treat	3	1 2 3

Number of Observations Read	15
Number of Observations Used	15

생성 환경: SAS 시스템('Local', X64_7PRO), 생성 일시: 2014년 03월 30일, 오후 12시01분51초

페이지 나누기
일원분산분석
결과
The ANOVA Procedure

Dependent Variable: y

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	3.41200000	1.70600000	4.60	0.0329
Error	12	4.45200000	0.37100000		
Corrected Total	14	7.86400000			

R-Square	Coeff Var	Root MSE	y Mean
0.433876	10.91573	0.609098	5.580000

Source	DF	Anova SS	Mean Square	F Value	Pr > F
treat	2	3.41200000	1.70600000	4.60	0.0329

생성 환경: SAS 시스템('Local', X64_7PRO), 생성 일시: 2014년 03월 30일, 오후 12시01분51초

해석?

The Analysis of Variance

● ANOVA 분석 후 작업

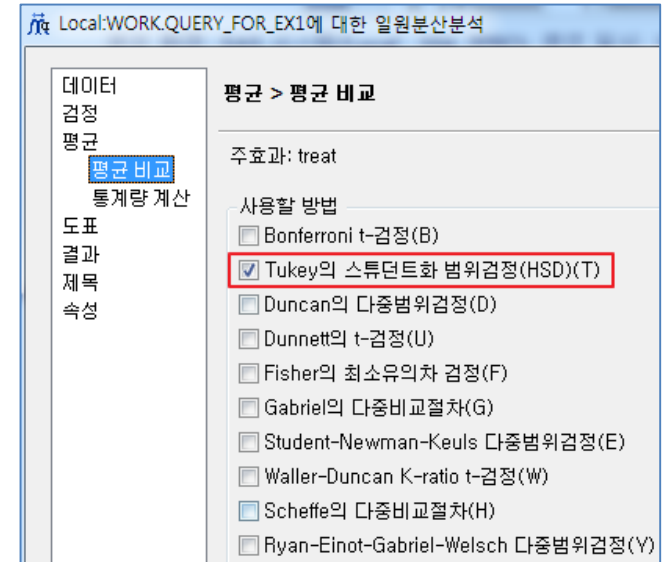
- 유의한 차이가 나왔으면 어떤 의미가 있는 지 궁금하지 않을까? 그게 바로 사후분석입니다. 통계 소프트웨어마다 사용 가능한 사후검정 방법이 다름.

Fixed Effect Model인 경우만 가능!!

각각의 방법에 대한 장단점이 있어서,
어떤 방법을 추천해줘야 할지 어려움.

사람들이 가장 많이 사용하는 방법은 Tukey, Duncan, Scheffe입니다.

저희 수업은 Tukey를 사용하겠습니다.



The Analysis of Variance

● Tukey 결과

The ANOVA Procedure

Tukey's Studentized Range (HSD) Test for y

Note: This test controls the Type I experimentwise error rate, but it generally has a higher Type II error rate than REGWQ.

Alpha	0.05
Error Degrees of Freedom	12
Error Mean Square	0.371
Critical Value of Studentized Range	3.77289
Minimum Significant Difference	1.0277

Means with the same letter are not significantly different.			
Tukey Grouping	Mean	N	treat
A	6.2000	5	3
A			
B	5.5000	5	1
B			
B	5.0400	5	2

- 최종 결론: 1(A)약의 경우 다른 약들과 차이는 나지 않고,
3(C)이 효과가 가장 크고, 2(B)는 효과가 가장 작다.

The Analysis of Variance(Model Checking)

- 오차의 가정 검토 – 잔차를 이용해서 확인

- ε_{ij} : Random error in the j-th observation on the i-th treatment,
following $NID(o, \sigma^2)$ with common variance σ^2 .

NID = Normally and Identically Distributed

=> 등분산성, 정규성, 독립성!!

The Analysis of Variance(Model Checking)

- 등분산성 Check(Levene's Test)

Local:WORK.QUERY_FOR_EX1에 대한 일원분산분석

데이터
검정
평균
평균 비교
통계량 계산
도표
결과
제목
속성

검정
☐ Welch의 분산 가중 일원분산분석(W)

등분산성 검정
☐ Bartlett의 검정(B)
☐ Brown Forsythe 검정(F)
☒ Levene의 검정(L)

The ANOVA Procedure

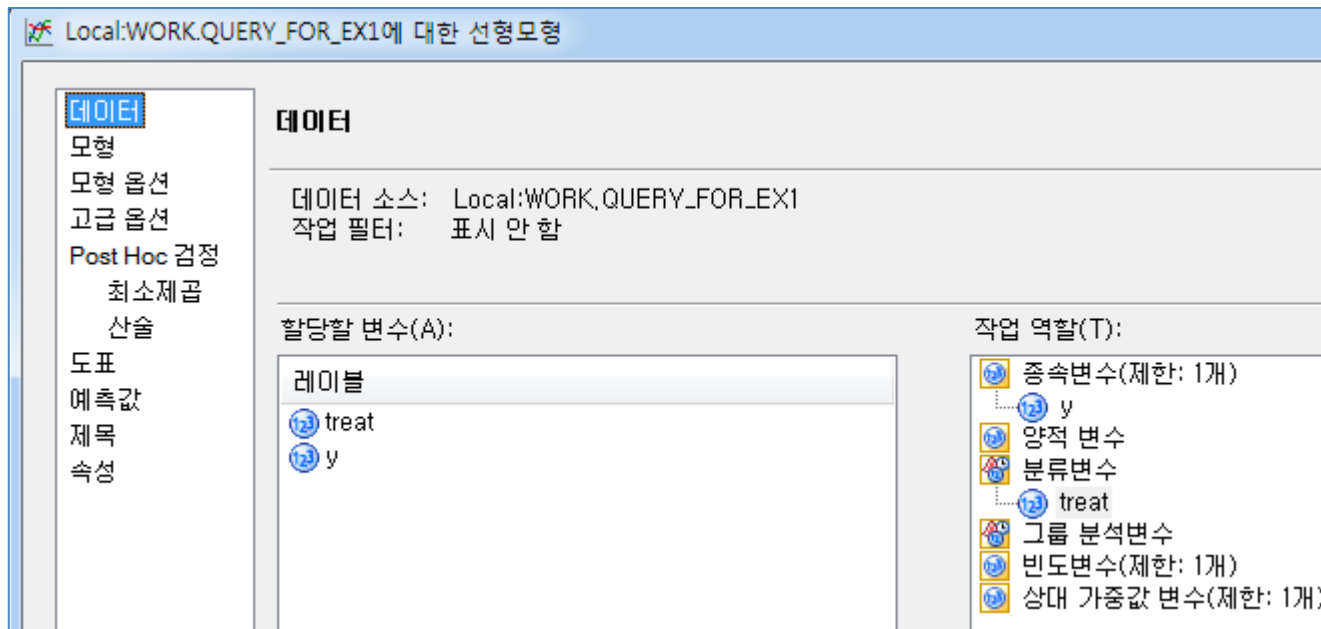
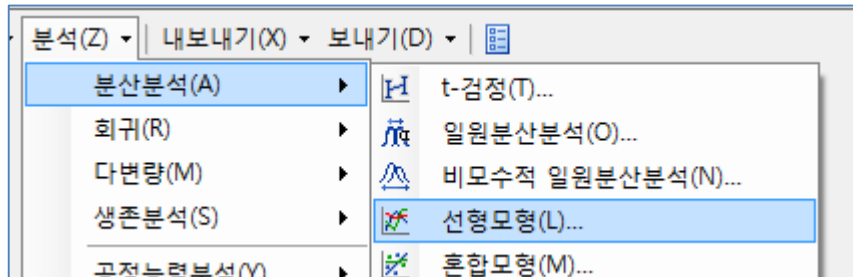
Levene's Test for Homogeneity of y Variance ANOVA of Squared Deviations from Group Means

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
treat	2	0.4645	0.2323	2.19	0.1545
Error	12	1.2724	0.1060		

- 등분산성 만족?

The Analysis of Variance(Model Checking)

- 정규성, 독립성, 등분산성 Check: 1단계



The Analysis of Variance(Model Checking)

- 정규성, 독립성, 등분산성 Check: 2단계

Local:WORK.QUERY_FOR_EX1에 대한 선형모형

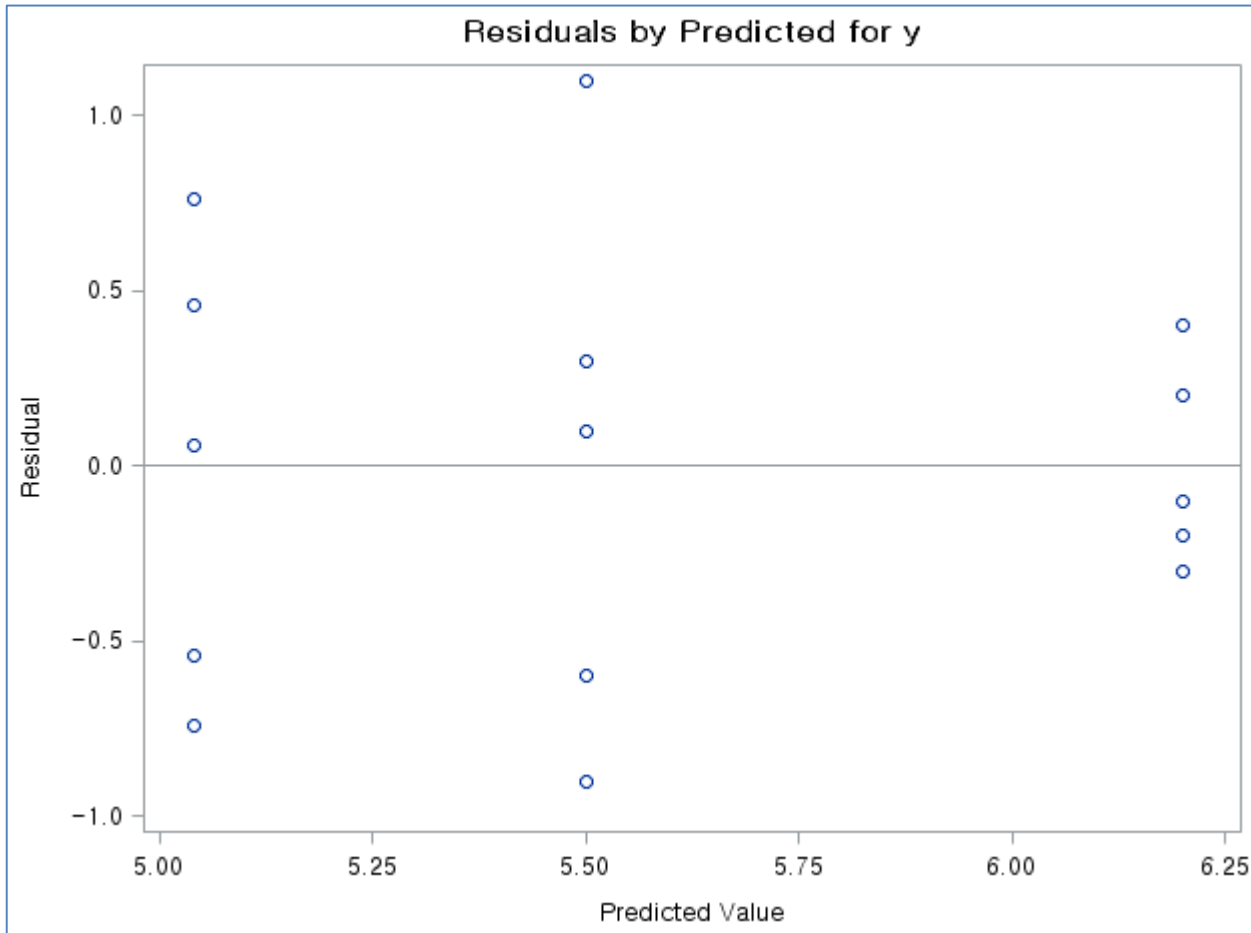
데이터	모형				
모형 모형 옵션 고급 옵션 Post Hoc 검정 최소제곱 산술 도표	<table border="1"><thead><tr><th>분류변수 및 양적변수(V):</th><th>효과(E):</th></tr></thead><tbody><tr><td> treat</td><td><div>주효과(M)</div><div>교호작용 효과(O)</div><div>지분효과(N)</div></td></tr></tbody></table>	분류변수 및 양적변수(V):	효과(E):	treat	<div>주효과(M)</div> <div>교호작용 효과(O)</div> <div>지분효과(N)</div>
분류변수 및 양적변수(V):	효과(E):				
treat	<div>주효과(M)</div> <div>교호작용 효과(O)</div> <div>지분효과(N)</div>				

Local:WORK.QUERY_FOR_EX1에 대한 선형모형

데이터	도표
데이터 모형 모형 옵션 고급 옵션 Post Hoc 검정 최소제곱 산술 도표 예측값 제목 속성	<div><input checked="" type="checkbox"/> 선형모형 분석에 대한 도표 표시(H)</div> <div><input type="radio"/> 현재 데이터 선택에 맞는 모든 도표(A)</div> <div><input checked="" type="radio"/> 사용자 정의 도표 리스트(U)</div> <div>사용자 정의 도표(P):<ul style="list-style-type: none"><input checked="" type="checkbox"/> 진단도표<input type="checkbox"/> 잔차도표<input type="checkbox"/> 상자도표</div>

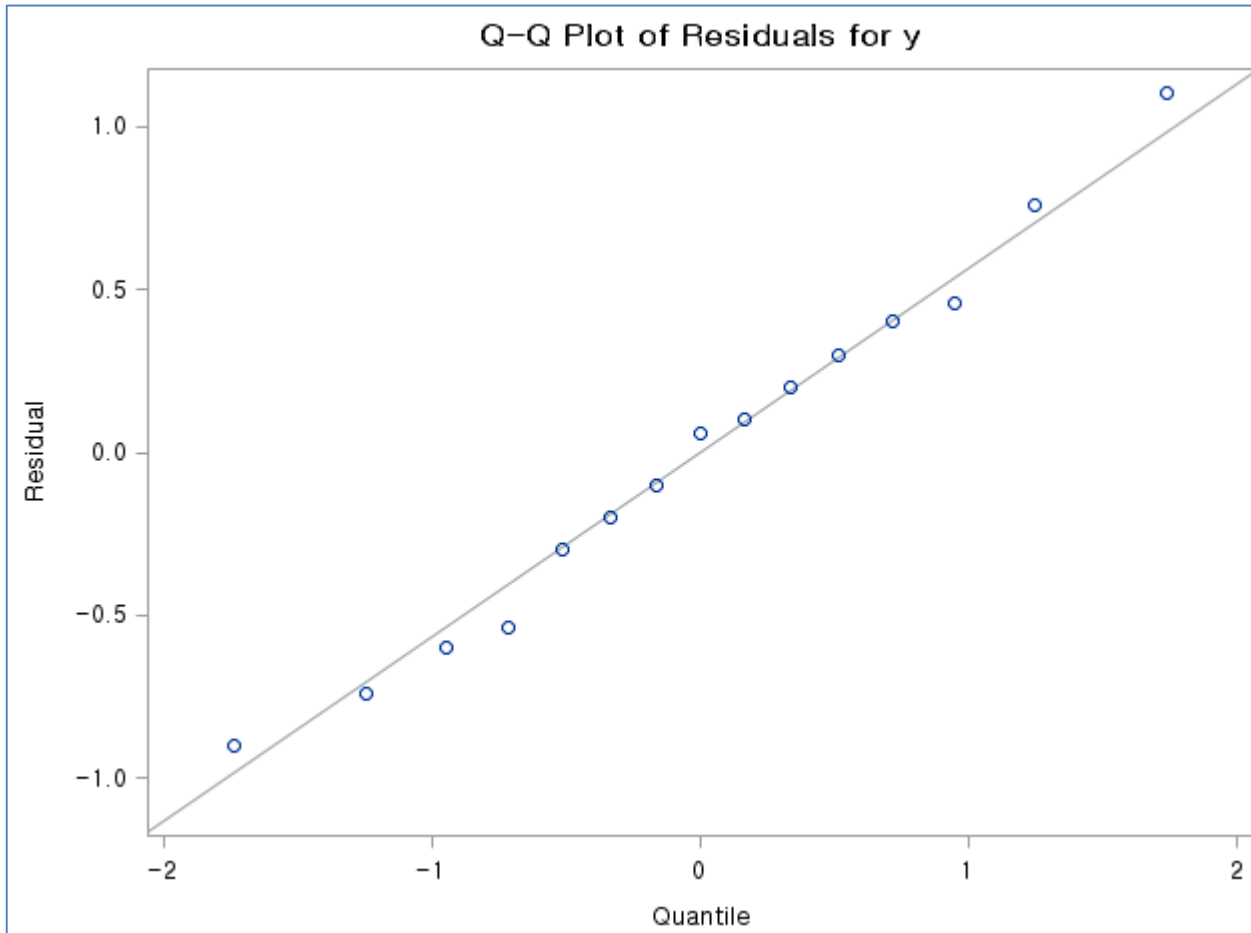
The Analysis of Variance(Model Checking)

- 정규성, 독립성, 등분산성 Check: 3-1단계 **Residual Plot**(등분산성, 독립성)



The Analysis of Variance(Model Checking)

- 정규성, 독립성, 등분산성 Check: 3-2단계 **Q-Q Plot(정규성)**



The Analysis of Variance(Model Checking)

- 선형모형에 그대로 적용 시 문제점

Dependent Variable: y

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	3.41200000	1.70600000	4.60	0.0329
Error	12	4.45200000	0.37100000		
Corrected Total	14	7.86400000			

R-Square	Coeff Var	Root MSE	y Mean
0.433876	10.91573	0.609098	5.580000

Source	DF	Type I SS	Mean Square	F Value	Pr > F
treat	2	3.41200000	1.70600000	4.60	0.0329

Source	DF	Type III SS	Mean Square	F Value	Pr > F
treat	2	3.41200000	1.70600000	4.60	0.0329

Parameter	Estimate		Standard Error	t Value	Pr > t
Intercept	6.200000000	B	0.27239677	22.76	<.0001
treat 1	-0.700000000	B	0.38522721	-1.82	0.0942
treat 2	-1.160000000	B	0.38522721	-3.01	0.0108
treat 3	0.000000000	B			

Note: The X'X matrix has been found to be singular, and a generalized inverse was used to solve the normal equations. Terms whose estimates are followed by the letter 'B' are not uniquely estimable.

- 분류변수로 설정된 변수를 가변수화 해서 적용.

The Analysis of Variance(Model Checking)

● 선형모형에 그대로 적용 시 문제점 - 해결책

	treat	y	treat_A	treat_B
1	1	4.6	1	0
2	1	5.8	1	0
3	1	5.6	1	0
4	1	4.9	1	0
5	1	6.6	1	0
6	2	4.3	0	1
7	2	5.8	0	1
8	2	5.1	0	1
9	2	5.5	0	1
10	2	4.5	0	1
11	3	6.4	0	0
12	3	6.6	0	0
13	3	5.9	0	0
14	3	6.0	0	0
15	3	6.1	0	0

작업 역할(T):	
종속변수(제한: 1개)	y
양적 변수	treat_A treat_B
분류변수	
그룹 분석변수	
빈도변수(제한: 1개)	
상대 가중값 변수(제한: 1개)	

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	3.41200000	1.70600000	4.60	0.0329
Error	12	4.45200000	0.37100000		
Corrected Total	14	7.86400000			

R-Square	Coeff Var	Root MSE	y Mean
0.433876	10.91573	0.609098	5.580000

Source	DF	Type I SS	Mean Square	F Value	Pr > F
treat_A	1	0.04800000	0.04800000	0.13	0.7253
treat_B	1	3.36400000	3.36400000	9.07	0.0108

Source	DF	Type III SS	Mean Square	F Value	Pr > F
treat_A	1	1.22500000	1.22500000	3.30	0.0942
treat_B	1	3.36400000	3.36400000	9.07	0.0108

Parameter	Estimate	Standard Error	t Value	Pr > t
Intercept	6.200000000	0.27239677	22.76	< .0001
treat_A	-0.700000000	0.38522721	-1.82	0.0942
treat_B	-1.160000000	0.38522721	-3.01	0.0108

- 가변수로 바뀐 treat_A와 treat_B는 양적 변수로 설정(IF문).
- 주의 메시지 없음.
- 결과는 동일하게 나오지만, 분류변수가 많아지면 해석하는데 있어서 어려움이 있으니 주의!!

The Analysis of Variance

- 직교대비의 적용(다이어트 약 예제 활용)

- 처리 1(A)과 처리 3(C)의 평균이 같은가?

$$H_0: \mu_1 = \mu_3$$

$$H_0: (1)\mu_1 + (0)\mu_2 + (-1)\mu_3 = 0$$

```
PROC glm DATA=EX1;  
  CLASS treat;  
  MODEL y = treat;  
  contrast 'treatment' treat 1 0 -1;  
RUN;
```

- 처리 1(A)과 처리 3(C)의 평균은 처리 2(B)의 평균과 같은가?

$$H_0: \frac{\mu_1 + \mu_3}{2} = \mu_2$$

$$H_0: (1)\mu_1 + (-2)\mu_2 + (1)\mu_3 = 0$$

```
PROC glm DATA=EX1;  
  CLASS treat;  
  MODEL y = treat;  
  contrast 'treatment' treat 1 -2 1;  
RUN;
```

여기서는 가변수가 아닌 **기존의 변수를** 넣어야 함!!