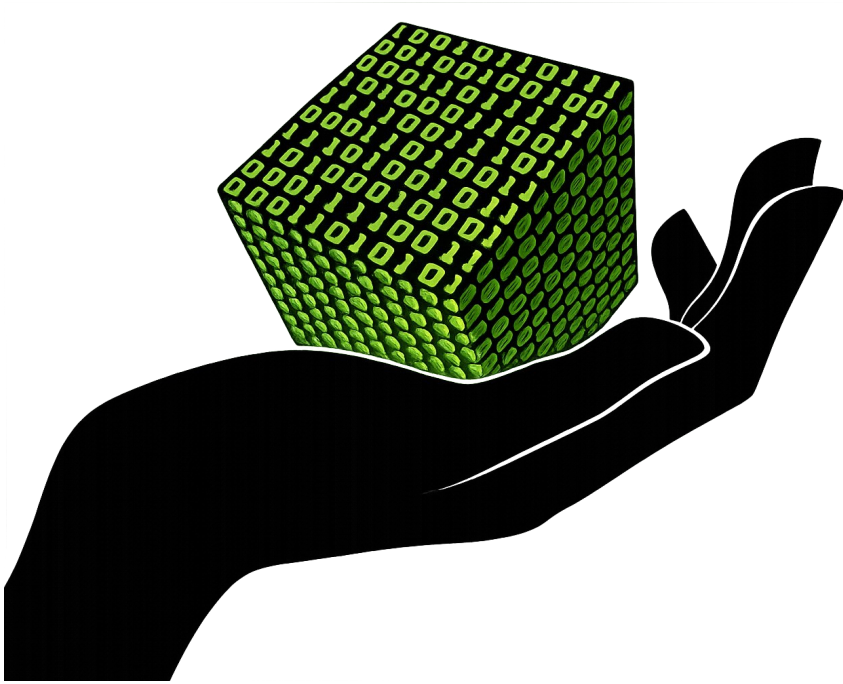




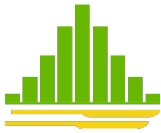
# **dBSTAT Tutorial**



# Data Management



*They had a large canvas bag, which tied at the mouse with strings.*



# dBSTAT

Data: Database.dbf

dBSTAT 에서 Database 를 생성하고 자료를 변환하는 방법은 SPSS 와 유사하다.

ID	AGE	SEX	BT	SERVICE	SEVERITY	DATE
1 H01	50	1	36.7	내과	low	2009.06.03
2 H02	23	2	36.8	외과	low	2009.05.28
3 H03	68	2	37.4	내과	low	2008.12.01
4 H04	22	1	37.3	산부인과	low	2008.02.28
5 H05	8	2	35.9	소아과	low	2008.12.28
6 H06	35	2	39.5	외과	high	2007.11.11
7 H07	80	1	38.3	내과	high	2007.04.30
8 H08	73	2	37.1	내과	low	2007.06.09
9 H09	13	1	37.8	소아과	middle	2007.05.31
10 H10	59	1	36.1	내과	middle	2007.05.16

번호	필드명	형태	길이	소수점
1	ID	C	3	0
2	AGE	N	2	0
3	SEX	C	1	0
4	BT	N	4	1
5	SERVICE	C	8	0
6	SEVERITY	C	6	0
7	DATE	D	8	0



DBSTAT 5 (교육용 버전)과 동영상 강의 파일은 다음 주소에서 내려받을 수 있다.

<http://www.webhard.co.kr> (ID: MedStat PW: 12345)

## 1 Database 생성

[새 파일] 버튼을 선택하거나 메뉴에서 [파일] ▶ [새 파일] 선택하면 나타나는 [새로운 DBF 파일구조] 창에서 다음과 같이 새로운 Database 를 만든다.

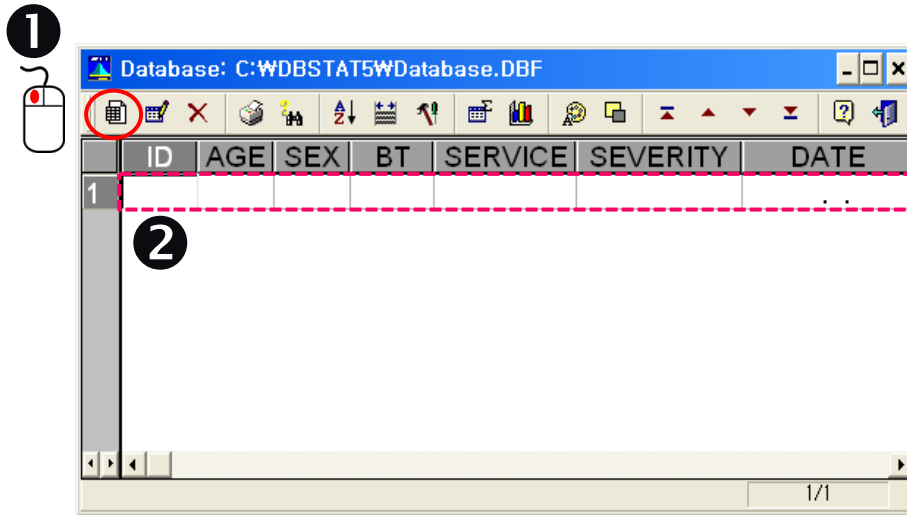
- ① [추가] 버튼을 클릭하여 새로운 변수를 만든다.
- ② [핀드] 이름, 형태, 길이, 소수점을 입력한다..
- ③ [생성] 버튼을 클릭하면 새 Database File 이 만들어진다.



## 필드(변수) 생성 규칙

1. [필드명]: 변수명을 입력한다. 변수명은 10 자 이내이며 자동으로 대문자로 입력된다.
2. [형 태]: 변수 형태 (C = 문자, N = 숫자, D = 날짜) 선택한다.
3. [길 이]: 입력자료의 소수점을 포함한 최대길이. 날짜형태는 8 자리수로 정해져 있다.
4. [소수점]: 소수점 이하 자리수이다. [예] BT = 36.8 → 길이 = 4, 소수점 = 1

## 새로운 레코드 생성



- ① [추가] 버튼을 클릭하여 새로운 레코드를 만든다.
- ② 변수의 자료를 입력한다.

셀에 자료 입력후에 [Enter] 키를 치면 커서가 다음 우측 변수로 이동하며 마지막 변수에서는 다음 레코드 첫번째 변수로 이동한다.

## 2 Data 변환

- dBSTAT 은 SPSS 와 동일하게 변수 단위로 자료가 변환된다.
- EXCEL 조건문에 사용되는 IF 조건함수와 사용법이 동일한 IF 함수를 이용하면 자료변환을 쉽게 할 수 있다.

새로운 변수 GRADE 를 만들고 SEVERITY = 'low' → GRADE = 1, SEVERITY = 'middle' → GRADE = 2, SEVERITY = 'high' → GRADE = 3 으로 변환하여 보자.

## 새로운 변수 생성

번호	필드명	형태	길이	소수점
1	ID	C	3	0
2	AGE	N	2	0
3	SEX	C	1	0
4	BT	N	4	1
5	SERVICE	C	8	0
6	SEVERITY	C	6	0
7	GRADE	N	1	0

	ID	AGE	SEX	BT	SERVICE	SEVERITY	GRADE
1	H01	50	1	36.7	내과	low	
2	H02	23	2	36.8	외과	low	
3	H03	68	2	37.4	내과	low	
4	H04	22	1	37.3	산부인과	low	
5	H05	8	2	35.9	소아과	low	
6	H06	35	2	39.5	외과	high	
7	H07	80	1	38.3	내과	high	
8	H08	73	2	37.1	내과	low	
9	H09	13	1	37.8	소아과	middle	
10	H10	59	1	36.1	내과	middle	

[필드] 버튼을 클릭하거나 [필드] ▶ [자료구조] 메뉴를 선택하면 [자료구조] 창이 보인다.

- ① [추가] 버튼을 클릭하여 새로운 변수 **GRADE** 를 만든다.
- ② [필드] 이름, 형태, 길이, 소수점을 입력한다..
- ③ [저장] 버튼을 클릭하면 자료구조를 변경하면 새 변수 GRADE 가 만들어진다.

## 변수값 변환

[변환] 버튼 클릭 또는 [자료] ▶ [자료 변환] 메뉴 선택 → [자료 변환] 상자가 나타난다.

① [치환필드] → 변수 GRADE 선택한다.

② [표현값] → [?] 버튼을 클릭한다.

③ [표현식] 직접 표현식을 입력하거나 필드 상자와 연산자 상자에서 항목을 선택하여 입력한다.  
[표현식] ⇨ IF( SEVERITY = 'low', 1, IF( SEVERITY = 'middle', 2, 3 ) )

- ① [치환필드] → 변수 GRADE 선택한다.
- ② [표현값] → [?] 버튼을 클릭한다.
- ③ [표현식] 직접 표현식을 입력하거나 필드 상자와 연산자 상자에서 항목을 선택하여 입력한다.  
[표현식] ⇨ IF( SEVERITY = 'low', 1, IF( SEVERITY = 'middle', 2, 3 ) )


④ [확인] 버튼을 클릭한다.

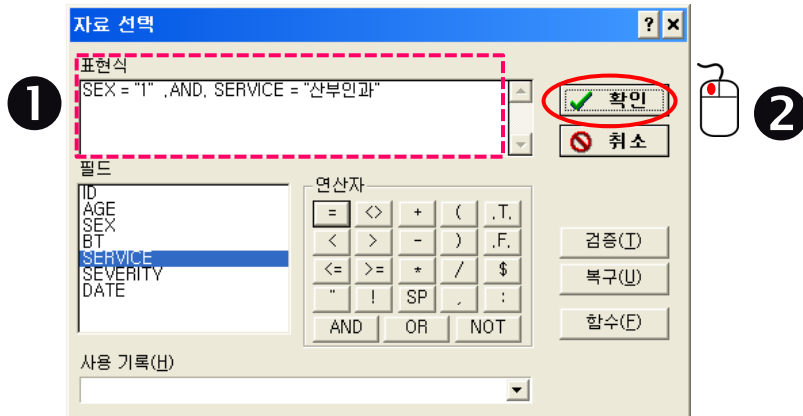
⑤ [자료 변환] 대화상자에서 [확인] 버튼을 클릭하여 작업을 마친다.

	ID	AGE	SEX	BT	SERVICE	SEVERITY	GRADE
1	H01	50	1	36.7	내과	low	1
2	H02	23	2	36.8	외과	low	1
3	H03	68	2	37.4	내과	low	1
4	H04	22	1	37.3	산부인과	low	1
5	H05	8	2	35.9	소아과	low	1
6	H06	35	2	39.5	외과	high	3
7	H07	80	1	38.3	내과	high	3
8	H08	73	2	37.1	내과	low	1
9	H09	13	1	37.8	소아과	middle	2

### 3 Data 선택

SPSS 에서는 조건문을 사용하여 원하는 자료를 검색하여 선택할 수 있다. 예제 Database Table 에서 SEX = '1' (남자) 이면서 SERVICE = '산부인과' (산부인과에 입원)인 case(레코드)가 있는지 검색하여 본다. 이러한 레코드가 검색되면 입력의 오류가 있음을 알 수 있다.

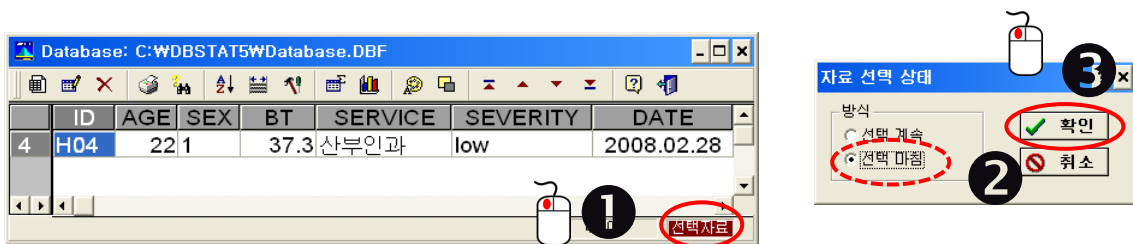
 [검색] 버튼을 클릭하거나 [자료] ► [자료 선택] 메뉴를 선택하면 [자료 선택] 대화상자가 나타난다. Data 변환에서의 [변환 표현식] 대화상자와 사용방법은 동일하다.



① [표현식] ⇨ SEX = '1' .AND. SERVICE = '산부인과' 입력한다.

② [확인] 버튼을 클릭하여 작업을 마친다.

조건문에 맞는 자료가 검색되어 보인다. 필터를 해제하려면 Database Table 하단에 [선택자료]를 클릭하여 나타나는 [자료 선택 상태] 상자에서 선택마침을 체크하고 [확인] 버튼을 클릭한다.



# Data File

EXCEL, SPSS, dBSTAT 에서 생성된 데이터 파일들은 서로 호환되거나 형식을 바꾸어 사용할 수 있다.

## 파일명

소프트웨어 파일은 각각 고유의 확장자를 가지고 있다. 확장자란 "database.dbf" 처럼 파일명을 점(.)으로 구분한 3 자리 문자이다.

 EXCEL ⇨ xls 또는 xlsx, SPSS ⇨ sav, dBSTAT ⇨ dbf 이다.



## 파일 사용

다른 프로그램의 데이터 파일을 사용하기 위해서는 메뉴 → [파일 열기]에서 직접 파일을 열거나 메뉴 → [다른 이름으로 저장]에서 다른 형식의 파일로 저장한 다음에 사용할 수 있다.

## Text 파일

첫 행에 변수명을 가진 데이터 파일을 일반 텍스트 파일로 저장하여야 한다.


### CSV (Comma Separated Value) 파일이란?

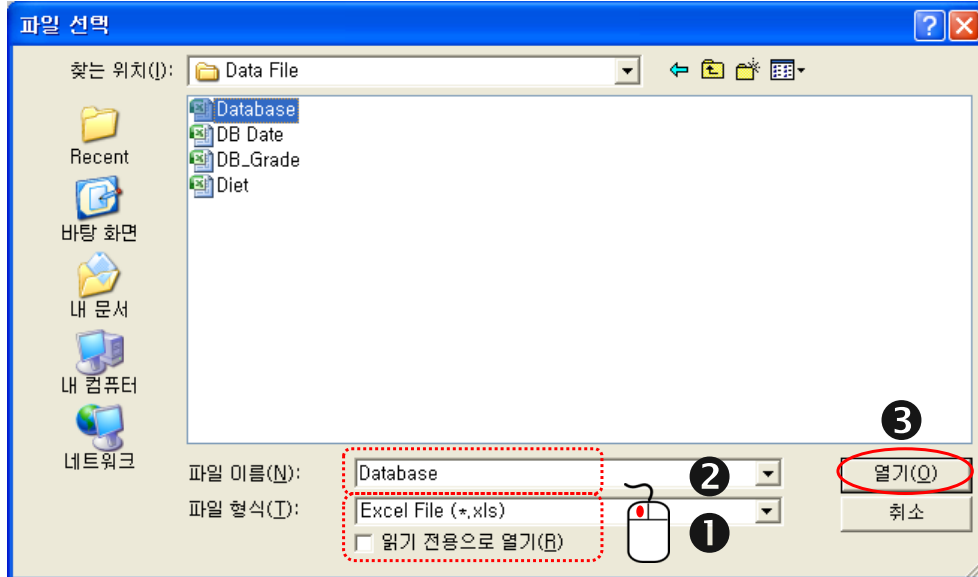
- ① 메모장에서 읽을 수 있는 서식이 없는 일반 텍스트(Text) 파일이다.
- ② 필드에 해당하는 변수들은 "," 쉼표로 구분된다.
- ③ 범주형 변수에서 숫자는 " (따옴표)로 입력하여야 한다. (예) '1'
- ④ 각 행마다 <Enter> 키로 끝난다.

ID	Age	Sex	BT	Service	Severity	Date
H01,	50,	'1',	36.7,	IM,	low,	2009-06-03
H02,	23,	'2',	36.8,	GS,	low,	2009-05-28
H04,	22,	'1',	37.3,	GYN,	low,	2008-02-28
H07,	80,	'1',	38.3,	IM,	high,	2007-04-30
H10,	59,	'1',	36.1,	PS,	middle,	2007-05-16

← 3 → | ← 2 → | ← 1 → | ← 4 → | ← 8 → | ← 6 → | ← 10 →


# Excel File

 [새 파일] 버튼을 클릭하거나 메뉴에서 [파일] ▶ [파일 열기] 선택한다.



- ① [파일 형식] ➡ Excel File (\*.xls) 형식을 선택한다.
- ② [파일 이름] ➡ Excel 파일명 (예, database.xls) 입력한다.
- ③ [열기] 버튼을 클릭하면 dBSTAT 에서 Excel 파일이 열린다.

	ID	AGE	SEX	BT	SERVICE	SEVERITY	DATE
1	H01	50	1	36.70	내과	low	2009-06-03
2	H02	23	2	36.80	외과	low	2009-05-28
3	H03	68	2	37.40	내과	low	2008-12-01
4	H04	22	1	37.30	산부인과	low	2008-02-28
5	H05	8	2	35.90	소아과	low	2008-12-28
6	H06	35	2	39.50	외과	high	2007-11-11
7	H07	80	1	38.30	내과	high	2007-04-30
8	H08	73	2	37.10	내과	low	2007-06-09
9	H09	13	1	37.80	소아과	middle	2007-05-31
10	H10	59	1	36.10	내과	middle	2007-05-16

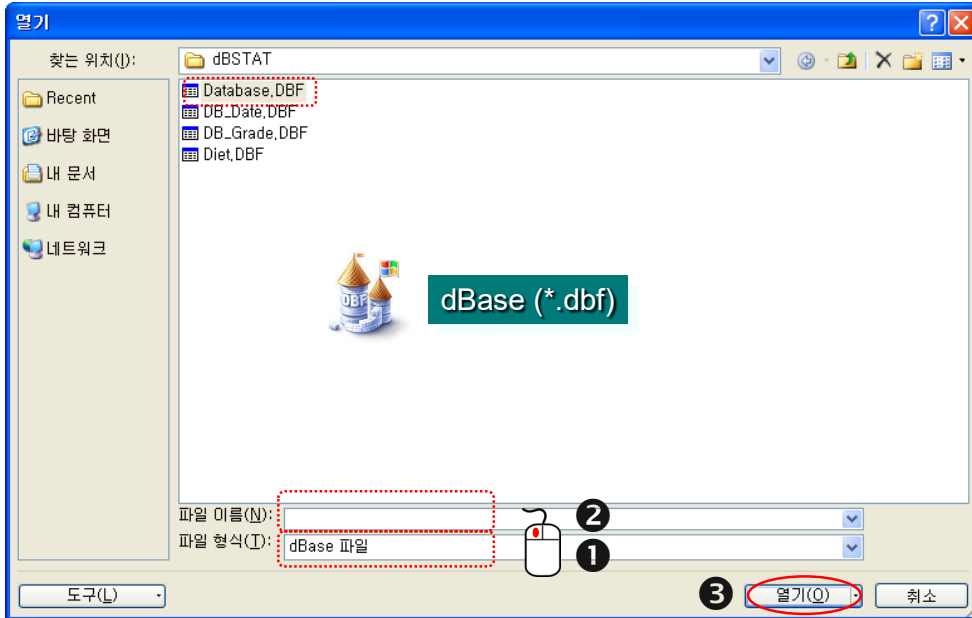
 dBSTAT 은 Excel 데이터 첫 행을 변수 이름으로 인식하여 자동적으로 읽어온다.  
Excel 통합 문서는 첫 번째 Worksheet 의 데이터만 읽을 수 있다.



# dBSTAT to Excel, SPSS

Data: Database.dbf

[파일] ▶ [열기] 선택하면 [파일 열기] 대화상자가 나타난다.



- ① [파일 형식] ➡ dBase (\*.dbf) 파일 형식을 선택한다.
- ② [파일 이름] ➡ dBSTAT 파일명 (예, Database.DBF) 선택 입력한다.
- ③ [열기] 버튼을 클릭하면 Excel 에서 dBSTAT 파일이 열린다.

	A	B	C	D	E	F	G	H
1	ID	AGE	SEX	BT	SERVICE	SEVERITY	DATE	
2	H01	50	1	36.7	내과	low	2009-06-03	
3	H02	23	2	36.8	외과	low	2009-05-28	
4	H03	68	2	37.4	내과	low	2008-12-01	
5	H04	22	1	37.3	산부인과	low	2008-02-28	
6	H05	8	2	35.9	소아과	low	2008-12-28	
7	H06	35	2	39.5	외과	high	2007-11-11	
8	H07	80	1	38.3	내과	high	2007-04-30	
9	H08	73	2	37.1	내과	low	2007-06-09	
10	H09	13	1	37.8	소아과	middle	2007-05-31	
11	H10	59	1	36.1	내과	middle	2007-05-16	
12								



Excel version 에 따라서는 [파일형식] 선택상자에 [dBase 파일] 형식이 안 보일 수 있다. 이 경우에는 [모든 파일] 형식을 선택하여 DBF 확장자(dbf)를 가지는 파일을 선택한다.

# 결측값 (missing value)

## 결측값 정의

- 자료가 누락되었거나 통계 분석에서 제외시키는 관측값을 결측값(분실값)이라고 한다.
- 문자형 변수는 공백인 경우 결측값으로 처리된다.
- **시스템 결측값**  
결측값은: Excel ⇨ 공백, dBSTAT ⇨ **0** (zero), SPSS ⇨ • (점)으로 기본 설정되어 있다.
- **사용자 결측값**  
기본 설정된 결측값을 실제 자료에 없는 특정한 숫자(예, -999)를 사용하여 바꿀 수 있다.

## 사용자 결측값:

Excel, SPSS, DBSTAT 에서 결측값은 다음 Table 에 설명되어 있다.

Software	문자형 변수	숫자형 변수		
		시스템 결측값		사용자 결측값
EXCEL	공백	공백	빈칸 표시	없음
DBSTAT	공백	<b>0</b>	빈칸 표시	제외시킬 특정한 숫자
SPSS	공백	공백	• (점) 표시	제외시킬 숫자, 특정 범위

※ dBSTAT 은 기본적으로 결측값이 0(zero)으로 설정되어 있으므로 0 을 통계분석에 포함시키려면 결측값을 Database 에서 사용되지 않는 특정 숫자 "-999" 등으로 변경하여야 한다.



다음 Database Table 에서 4 번째 Service = '산부인과'와 5 번째 Age = 8 를 통계 분석에서 제외시키도록 한다.

ID	Age	Sex	BT	Service	Severity	Date
H01	50	1	36.7	내과	low	2009-06-03
H02	23	2	36.8	외과	low	2009-05-28
H03	68	2	37.4	내과	low	2008-12-01
H04	22	1	37.3	산부인과	low	2008-02-28
H05	8	2	35.9	소아과	low	2008-12-28
H06	35	2	39.5	외과	high	2007-11-11
H07	80	1	38.3	내과	high	2007-04-30
H08	73	2	37.1	내과	low	2007-06-09
H09	13	1	37.8	소아과	middle	2007-05-31
H10	59	1	36.1	내과	middle	2007-05-16



# dBSTAT

Data: Database.dbf

## 1 시스템 결측값

ID	AGE	SEX	BT	SERVICE	SEVERITY	DATE	
1	H01	50	1	36.70	내과	low	2009-06-03
2	H02	23	2	36.80	외과	low	2009-05-28
3	H03	68	2	37.40	내과	low	2008-12-01
4	H04	22	1	37.30		low	2008-02-28
5	H05		1	35.90	소아과	low	2008-12-28
6	H06	35	2	39.50	외과	high	2007-11-11
7	H07	80	1	38.30	내과	high	2007-04-30

- ① 숫자형 변수 [Age] → record 5 셀을 0으로 만들면 시스템 결측값이 된다.
- ② 문자형 변수 [Service] → record 4 셀의 입력된 값을 삭제하면 결측값이 된다.

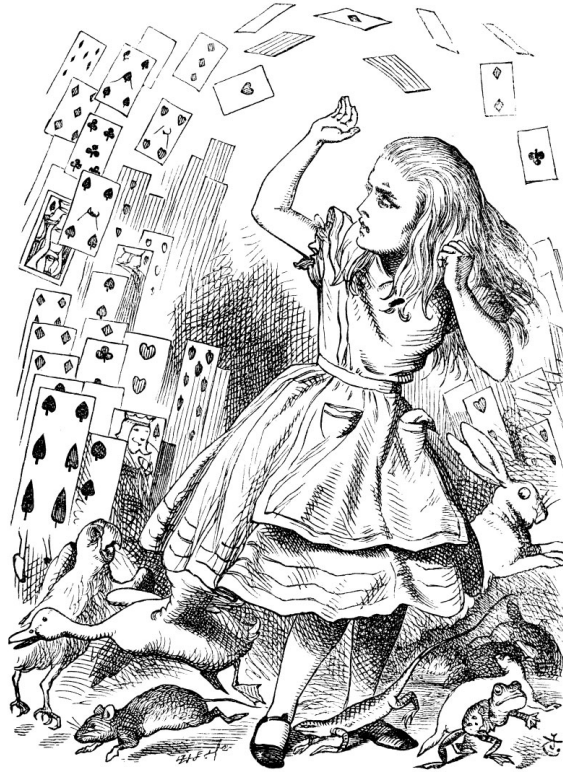
※ **Del** 키를 누르면 전체 record가 삭제되므로 주의를 요한다.

## 2 사용자 정의 결측값

dBSTAT에서 기본 설정된 결측값은 0이다. 통계분석에 0을 포함하여야 하는 경우에는 결측값을 별도로 정한다. 일반적으로 측정되지 않는 수(예, -9, 999 등)를 사용자 결측값으로 정한다.

- ① **[통계] ▶ [결측값]** 메뉴를 선택하면 **[결측값 설정]** 대화상자가 나타난다.
- ② dBSTAT에서 기본 설정된 결측값 0 → -9로 바꾼 다음 **[확인]**을 클릭한다.
- ③ DatabaseTable에서 S번째 record에서 **[AGE]** 0 → -9로 바꾼다.
- ④ 변수 **[Age]** 5번째 record 셀의 관측값(-9)이 결측값으로 처리되어 공백으로 보인다.

# Data Summary



*“I seem to see some meaning in them, after all.”*

dBSTAT 은 DataBase Statistics(데이터베이스 통계) 또는 The BioStatistics(생물통계)의 약자를 뜻하며 의학 생물학 분야의 전문 통계소프트웨어로 사용이 매우 편리하며 통계결과에 대한 해석을 하여 주므로 통계 초보자들이 사용하기에 적합한 소프트웨어이다.

## dBSTAT 특징

### Database

- dBSTAT 에서 데이터를 처리하는 방법은 EXCEL 과 비슷하여 사용법이 매우 용이하다.
- EXCEL 과 SPSS 파일을 읽을 수 있으며 dBSTAT 파일을 EXCEL, SPSS 에서 사용할 수 있다.

### Statistics

- **기술통계:** 평균, 표준편차, 중앙값, 사분위범위, 범위, 빈도분석, 정규분포, 이상값 검정 등.
- **비교통계:** t-test, ANOVA, Mann-Whitney test, Wilcoxon signed rank test, Sign test, Kruskal-Wallis test, Jonckheere-Terpstra, Friedman test, ANCOVA, Repeated measures ANOVA 등.
- **관계분석:** 상관분석, 단순회귀분석, 다중회귀분석, 판별분석, 로지스틱 회귀분석 등.
- **생존분석:** Kaplan-Meier method, Life table method, Log-rank test, Cox regression 등.
- **상호관계:** 군집분석, 주성분분석, 요인분석 등.
- **진단통계:** 민감도, 특이도, 예측도, ROC, Kappa test, 급내상관, Bland-Altman method 등.
- **표본크기:** 평균, 비율, 상관분석, 생존분석, 진단통계, 우월성, 비열등성, 동등성 검정 등.
- **교차분할표:** Chi-square test, McNemar test, odds ratio, risk ratio, likelihood ratio 등.



**통계마법사** 다른 통계 소프트웨어에서는 찾아볼 수 없는 독특한 기능이다. 통계분석 방법을 모르는 경우에 자동으로 가장 적합한 통계분석 방법을 찾아서 통계분석하고 통계결과를 해석해준다. 통계 초보자에게 매우 유용한 기능이다.

### Graphics

- 통계분석시 자동적으로 Excel 그래프가 생성되며 Word 또는 PowerPoint 에서 편집할 수 있다.
- 그래프 마법사가 있어 원하는 그래프 형태를 선택하면 간단히 그래프를 만들 수 있다.

dBSTAT 에서 작성할 수 있는 그래프의 종류는 크게 다음과 같다.

선그림, 면적그림, 원그림, 산점도, 상한-하한 그림, 오차막대, 상자그림, 파레토, 시계열그림, Polar 그림, Bubble, Gantt, Candlestick 그림 등.



## Exercise

### Data Summary



다음은 성인 25 명에서 혈액형을 조사한 자료이다.

ID	Blood	ID	Blood	ID	Blood
001	A	011	AB	021	O
002	B	012	A	022	O
003	B	013	O	023	AB
004	AB	014	B	024	B
005	O	015	B	025	A
006	O	016	O		
007	O	017	A		
008	B	018	O		
009	AB	019	A		
010	B	020	O		

1. 혈액형의 빈도를 구하시오.
2. 혈액형의 빈도를 막대그래프로 표현하시오.

#### [해설]

예제 자료에 대한 풀이를 dBSTAT 을 사용하여 시행한다.

예제 파일은 다음 웹사이트에서 다운로드 받을 수 있다.

 <http://www.webhard.co.kr> (ID: MedStat PW: 12345)



**dBSTAT**

Data: newborn.dbf

## 기술통계

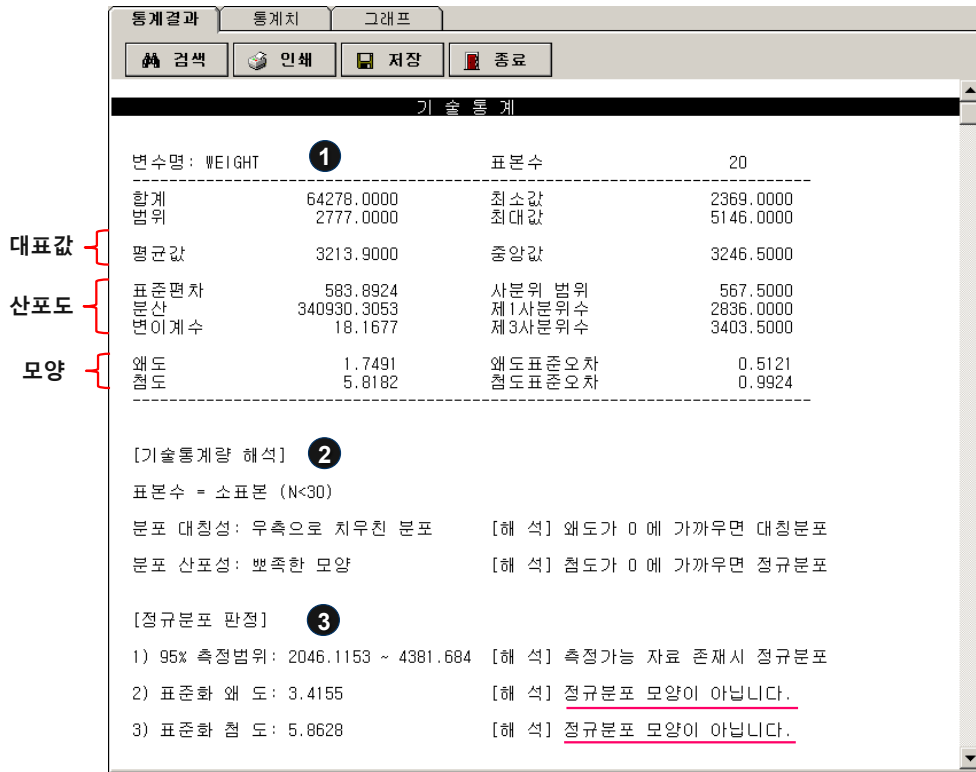
1 메뉴 → [통계] ▶ [기술통계] ▶ [마법사] → [자료변수] 메뉴를 선택한다.

[변수 선택] 대화상자가 나타난다.

The screenshot shows the dBSTAT software interface. On the left, a menu tree is visible with '통계(S)' expanded to '기술통계(D)', then '마법사(W)', and finally '자료변수(D)'. A green arrow points from the '자료변수(D)' menu item to the '변수 선택' dialog box. The dialog box has a title bar with a '2' and a mouse cursor icon. Inside, the '선택범위' section has '일부' selected. The variable 'WEIGHT' is highlighted in a list. On the right side of the dialog, the '확인' button is circled in red. Other buttons include '취소' and '제거'. A cartoon wizard character is visible in the top right corner of the dialog area.

① [변수 선택] 대화상자에서 변수 weight 를 선택한다.

② [확인] 버튼을 클릭하여 선택 작업을 마치면 기술통계 분석 결과가 출력된다.

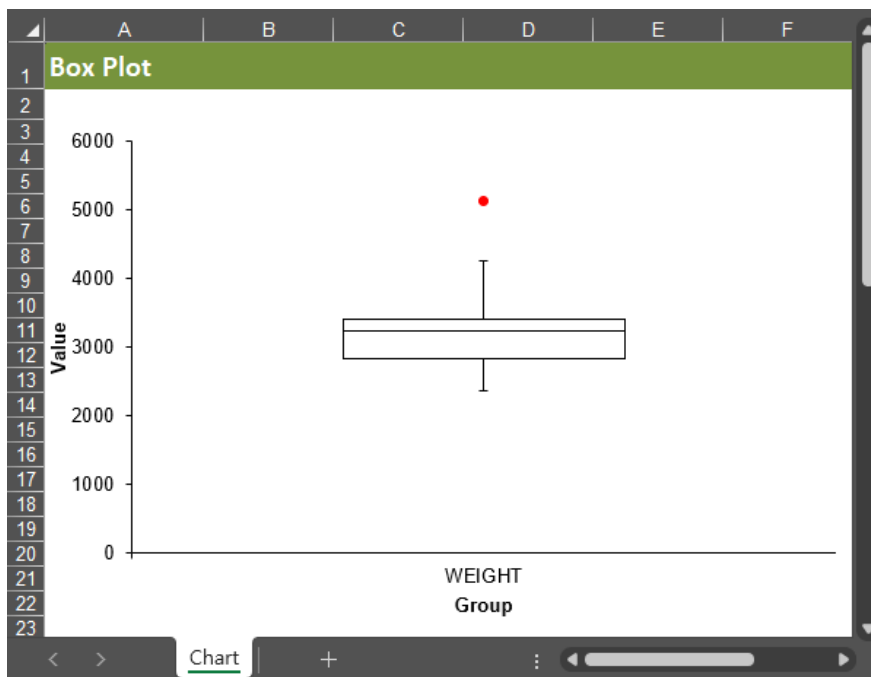


✎ [기술통계량 해석] 왜도 = 1.7 > 1, 첨도 = 5.8 > 3 ⇒ 우측으로 기운 뾰족한 분포

✎ [정규분포 판정] 표준화 왜도 = 3.41 > 1.96, 표준화 첨도 = 5.86 > 1.96 ⇒ 비정규분포

## Excel Chart

**Box-and-Whisker Plot:** Excel 창에 box-plot 이 나타난다.

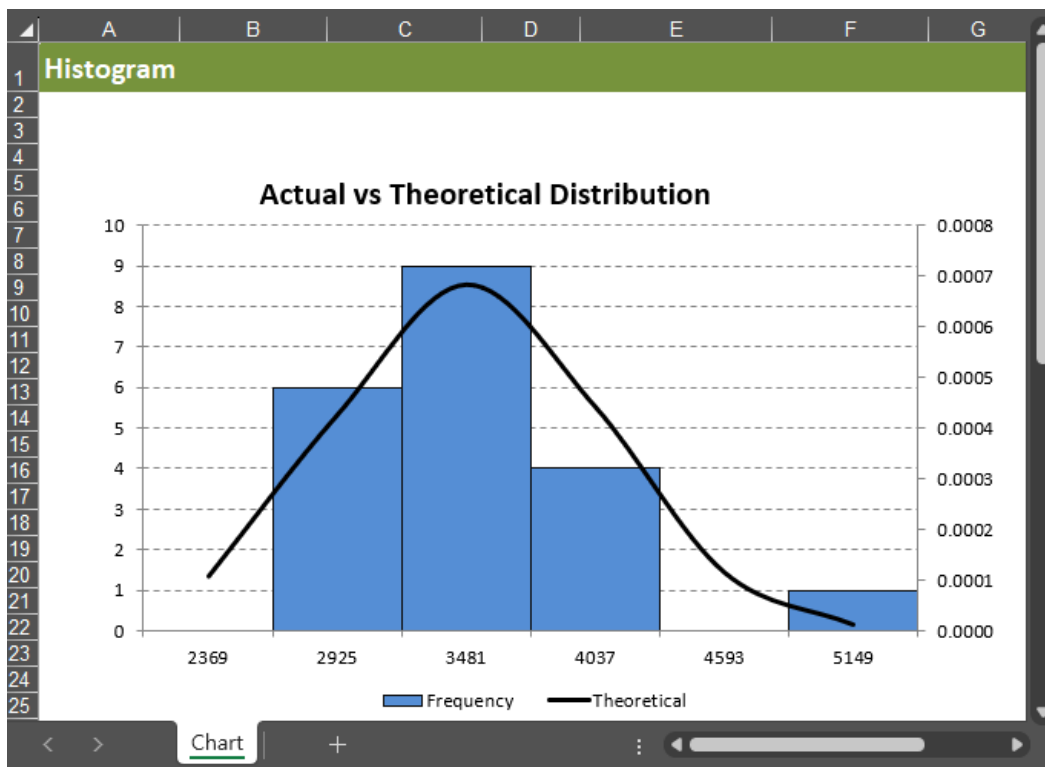


## 2 Histogram

[통계] ▶ [빈도분석] ▶ [도수분포표] 메뉴를 선택한다.

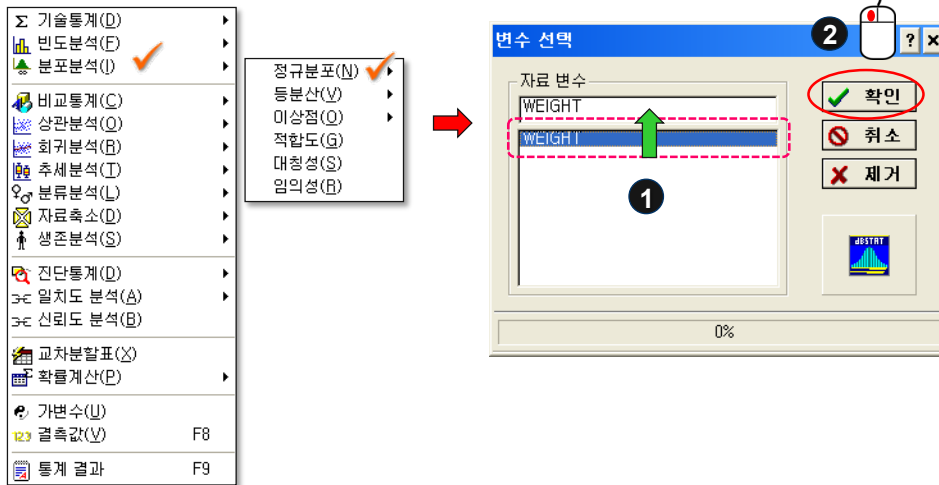
- ① [변수명] 상자에서 변수 weight 를 선택한다.
- ② [설정] 버튼 클릭 ➡ [계급 설정] 상자에 [계급수], [계급폭], [하한값]이 자동 입력된다.
- ③ [확인] 버튼을 클릭하여 선택 작업을 마치면 Histogram 이 출력된다.

## Excel Chart

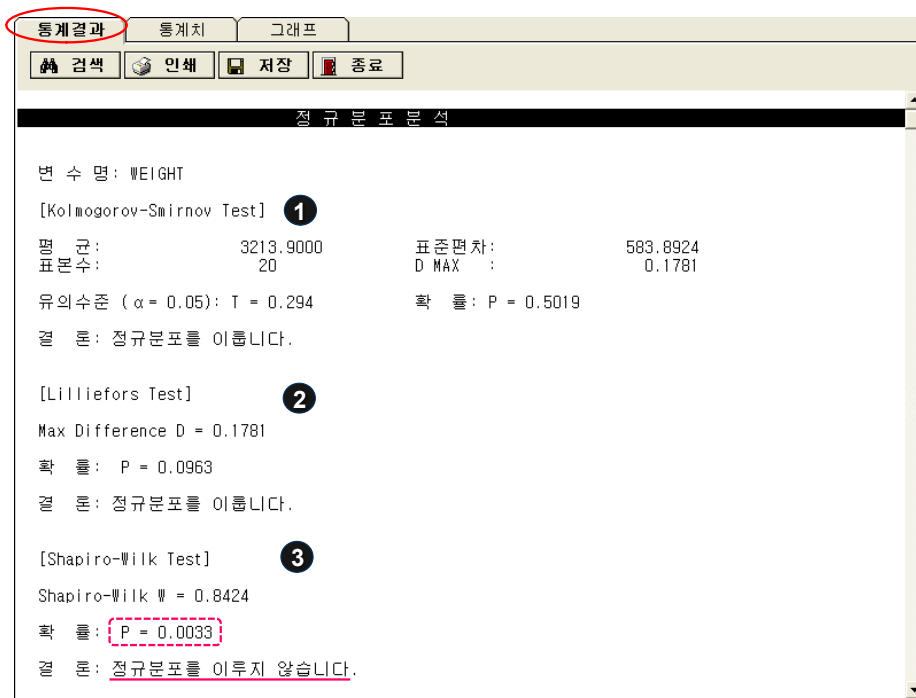


### 3 정규성 검정

[통계] ▶ [분포분석] ▶ [정규분포] → [자료변수] 메뉴를 선택한다.



- ① [변수 선택] 대화상자에서 변수 weight 를 선택한다.
- ② [확인] 버튼을 클릭하여 선택 작업을 마치면 정규분포 분석 결과가 출력된다.



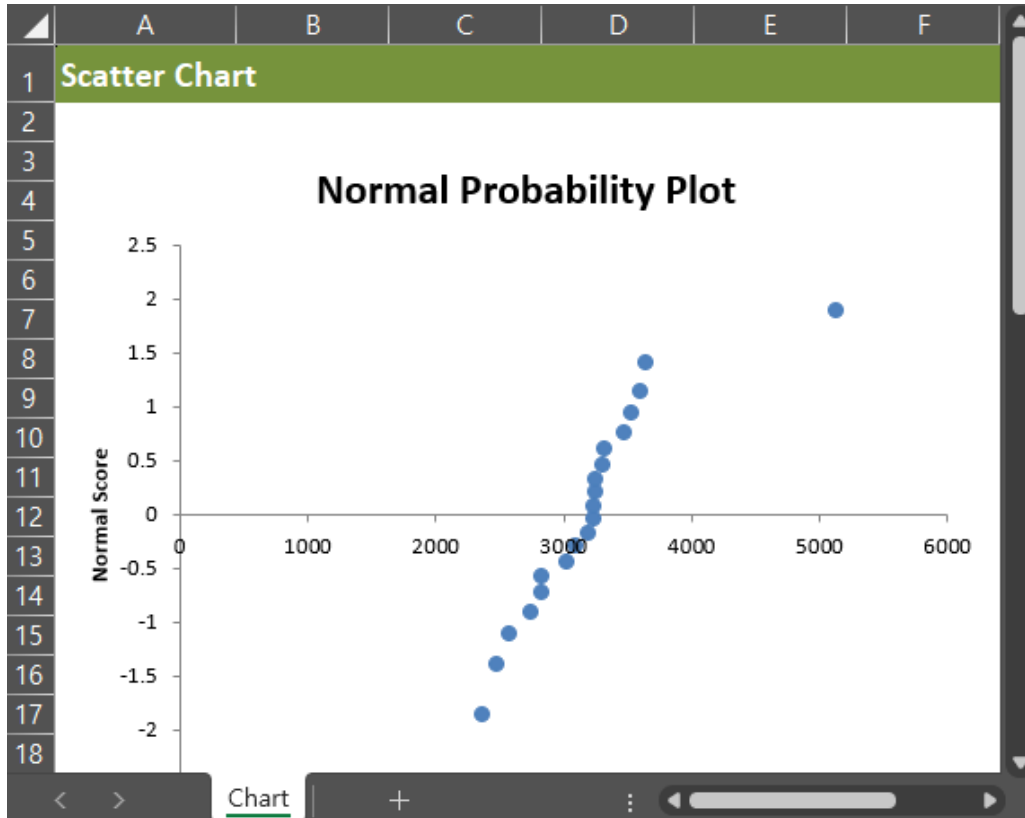
- ① **Kolmogorov-Smirnov** 검정:  $p = 0.50 > 0.05 \Rightarrow$  정규분포를 이룬다.
- ② **Lilliefors** 검정:  $p = 0.096 > 0.05 \Rightarrow$  정규분포를 이룬다.
- ③ **Shapiro-Wilk** 검정:  $P = 0.003 < 0.05 \Rightarrow$  정규분포를 이루지 않는다.

✎ 3 가지 검정법 중에서 Shapiro-Wilk 검정 결과만 비정규분포를 보인다.

✎ 소표본( $N < 30$ )이므로 모든 검정에서 정규성을 만족하지 못하면 비정규분포로 판정한다.

## Excel Chart

**Q-Q Plot:** Excel 창에서 Q-Q plot 이 나타난다.



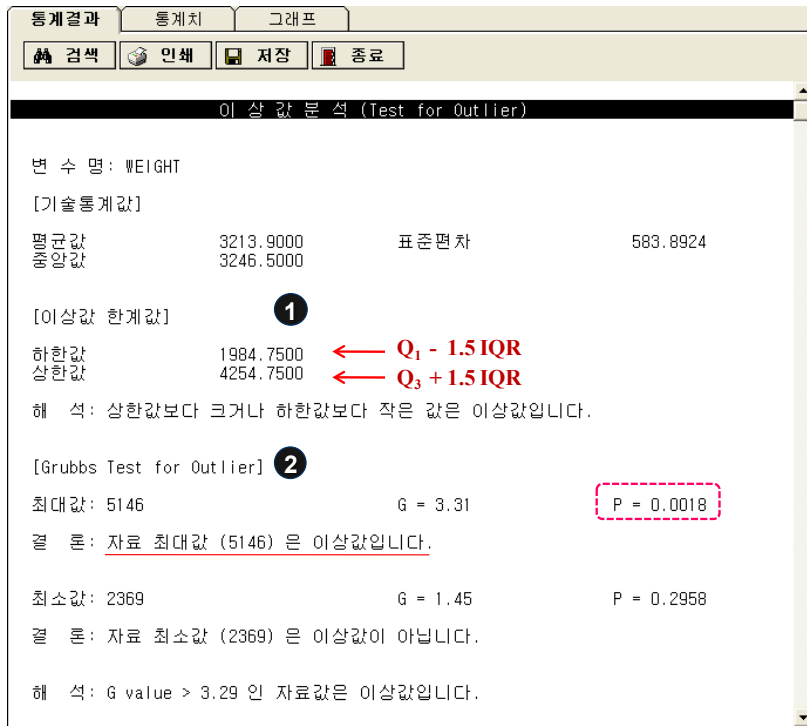
그래프 우측 상단에 이상값(outlier) (>5000 gm)가 보인다.

## 4 이상값 검정

### 1. 이상값 검정

[통계] ▶ [분포분석] ▶ [이상값] 메뉴를 선택한다.

[변수 선택] 대화상자에서 변수 weight 를 선택한 다음 [확인] 버튼을 클릭하여 선택 작업을 마치면 분석 결과가 출력된다.

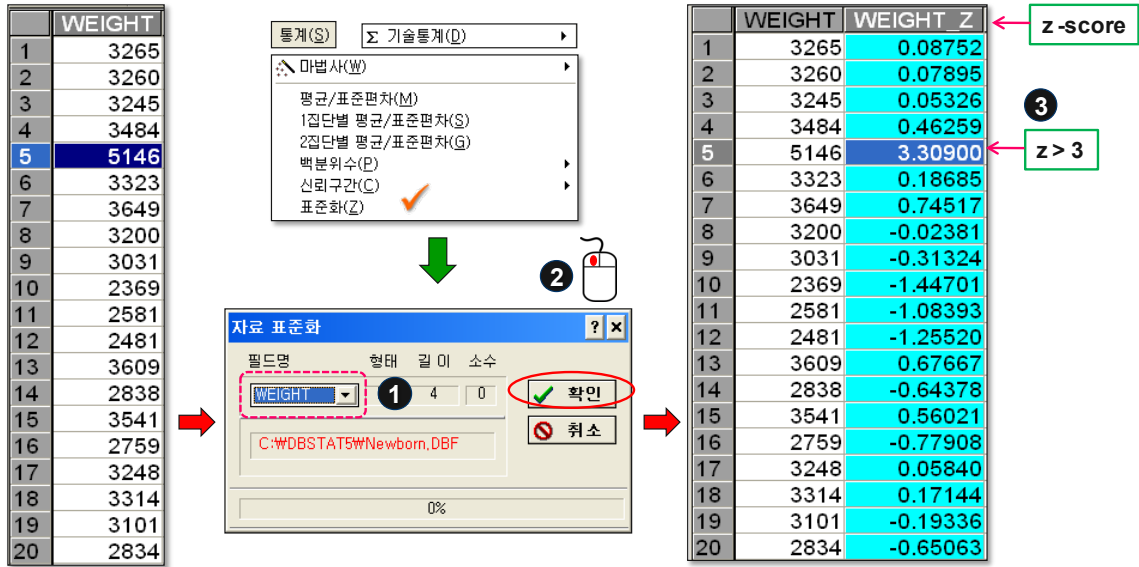


- ① [이상값의 한계값] 사분위수와 사분위범위에 의해 하한값과 상한값을 계산한다.
- ② [Grubbs test] 최대값 5146 ⇒  $p = 0.0018 < 0.05$  이므로 이상값으로 판정한다.

### 2. z-score

[통계] ▶ [기술통계] ▶ [표준화] 메뉴를 선택한다.

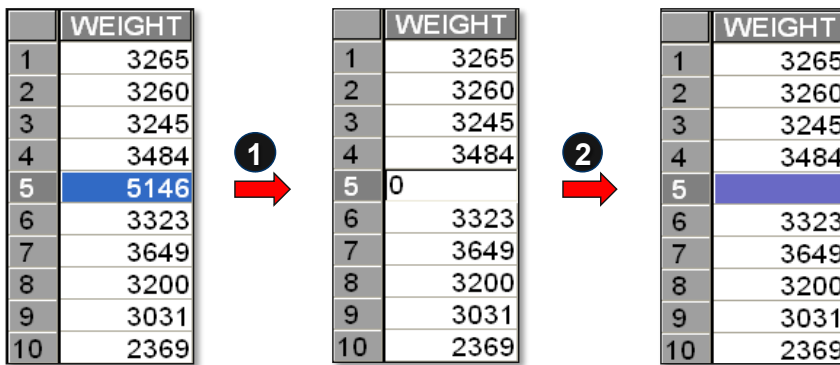
- ① [자료 표준화] 대화상자에서 변수 weight 를 선택한다.
- ② [확인] 버튼을 클릭하면 z-score 값을 가진 새로운 변수 WEIGHT\_Z 가 생성된다.



✎ 5 번째 record(5146)의 z-score = 3.3 > 3 이므로 이상값으로 판정한다.

## 이상값 제외

Database Table 에서 셀의 입력값을 0 으로 바꾸면 결측값으로 처리하여 분석에서 제외된다.



① 5 번째 record 5146 을 0 으로 만든다.

② 0 ⇨ 공백으로 변환되어 시스템 결측값으로 바뀐다.

## 5 결측값 설정

[통계] ▶ [결측값] 메뉴를 선택하면 [결측값 설정] 대화상자가 나타난다.

The process is illustrated in four steps:

- Step 1:** The 'Statistics' menu is open, and 'Missing Values' is selected.
- Step 2:** The 'Missing Values' dialog box is shown with the default value '0,0000'.
- Step 3:** The dialog box is shown with the value changed to '-999'.
- Step 4:** Two 'WEIGHT' tables are shown:
 

	WEIGHT
1	3265
2	3260
3	3245
4	3484
5	0
6	3323
7	3649
8	3200
9	3031
10	2369

 and
 

	WEIGHT
1	3265
2	3260
3	3245
4	3484
5	-999
6	3323
7	3649
8	3200
9	3031
10	2369


- ① dBSTAT 에서 기본 설정된 결측값은 0 이다. 결측값은 공백으로 처리되어 분석에서 제외된다.
- ② 0 을 분석에 포함하기 위해서는 결측값을 바꾸어야 한다. (예, -999)

이상값을 통계분석에서 제외한 후 정규분포 검정을 시행한 결과는 다음과 같다.

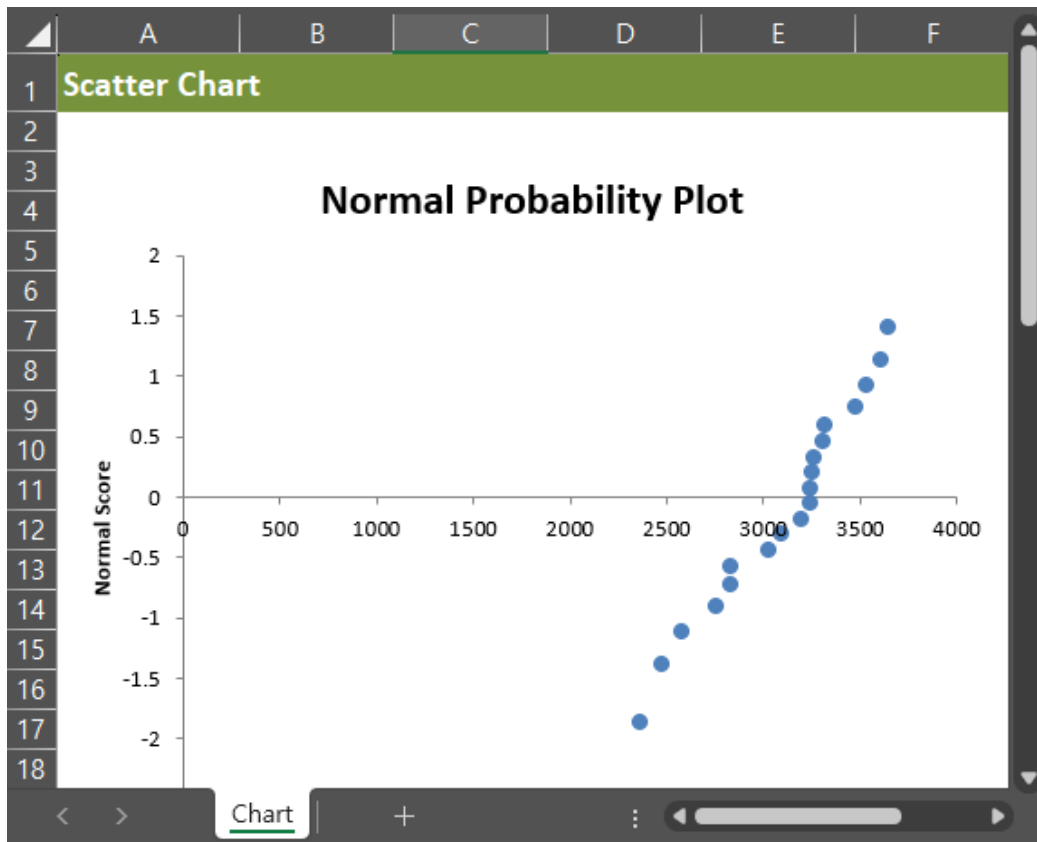
The screenshot shows the '통계결과' (Statistics Results) window in SPSS. The title bar reads '통계결과'. Below the title bar are buttons for '검색' (Search), '인쇄' (Print), '저장' (Save), and '종료' (End). The main content area is titled '정규분포분석' (Normality Test). The variable being tested is '변수명: WEIGHT'. The tests performed are the Kolmogorov-Smirnov Test, Lilliefors Test, and Shapiro-Wilk Test. The results for each test are as follows:

Test Name	Statistic	P-value	Conclusion
Kolmogorov-Smirnov Test	D MAX	0.1712	정규분포를 이룹니다.
Lilliefors Test	Max Difference D	0.100*	정규분포를 이룹니다.
Shapiro-Wilk Test	Shapiro-Wilk W	0.2782	정규분포를 이룹니다.

The Shapiro-Wilk test result 'P = 0.2782' is circled in red in the original image. The conclusion for the Shapiro-Wilk test is underlined in red.

 Shapiro-Wilks 검정:  $p = 0.278 > 0.05$  이므로 정규분포로 판정한다.

## Excel Chart



✎ Q-Q plot : 대부분 관측값들이 일직선을 이루게 되어 정규분포에 가깝다.

# Comparison



*“Once,” said the Mock Turtle at last, with a deep sigh,*

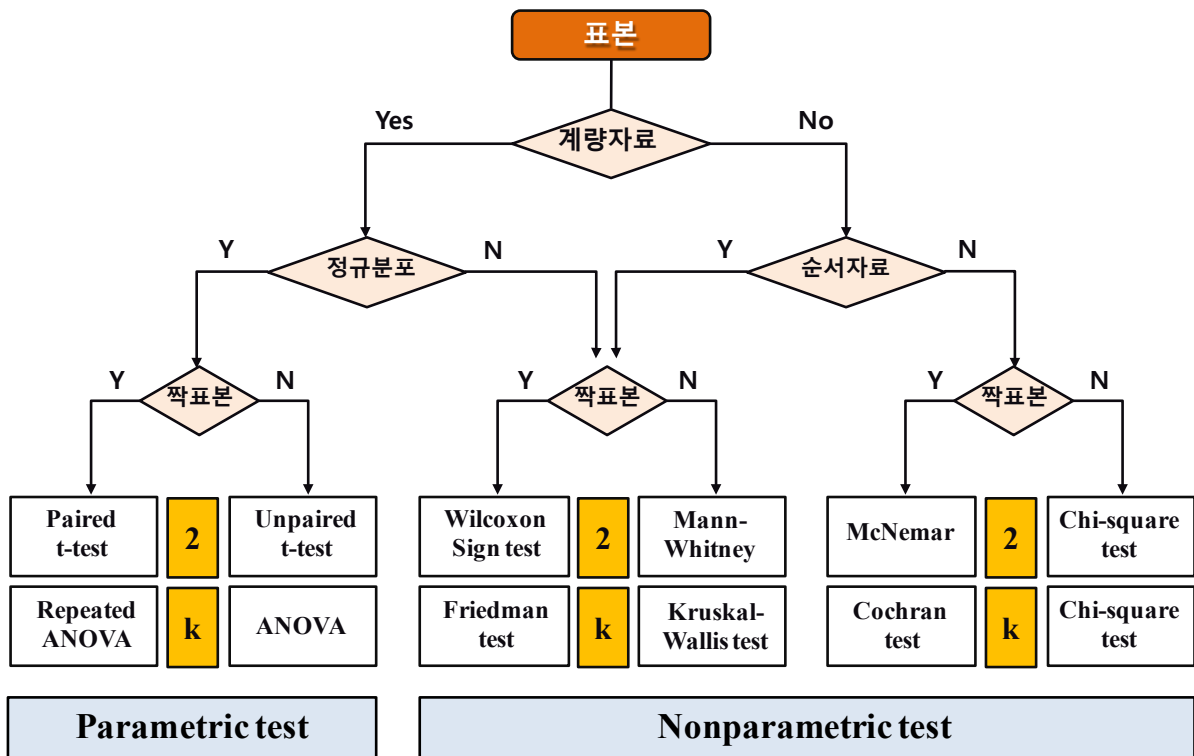
*“I was a real Turtle.”*

*Alice was thoroughly puzzled.*

# 비교통계 (Comparison)

의학연구에서 실험군과 대조군 집단들 간의 결과를 비교하는 연구가 가장 많이 시행되고 있다.

비교분석이란 표본 집단들 간에 표본통계량(평균, 분산, 비율)을 비교하는 통계분석을 뜻한다.



비교 통계 방법은 크게 다음과 같은 기준에 의하여 구분할 수 있다.

## 1. 자료형태

연속형 계량자료와 비연속형인 순서자료, 명목자료에 따라서 통계 검정법이 달라진다.

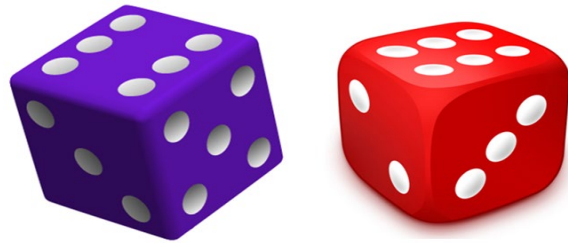
## 2. 정규성

표본 자료가 정규분포를 이루면 모수적 검정, 비정규분포이면 비모수적 검정을 선택한다.

## 3. 표본(집단) 수

비교하려는 표본(집단)의 수에 따라 통계분석 방법이 달라진다.

# Two Sample





## Exercise

### Student $t$ -test



다음은 심전도(EKG) 검사에서 정상 10 명과 비정상 10 명에서 혈중 콜레스테롤(단위: mg/dL)을 측정한 자료이다. 비정상 심전도 집단의 평균 콜레스테롤은 정상에 비하여 차이가 있는지 검정하시오.

Cholesterol			
정상 EKG		비정상 EKG	
141	136	202	293
189	186	256	237
173	177	283	249
198	124	194	223
204	136	184	271

#### [통계소프트웨어 사용]

##### 1. 통계적 검정법

- ① 표본평균 검정
- ② 독립표본, 표본 집단수 = 2  $\Rightarrow$  **independent sample t-test** 를 선택한다.

##### 2. 전제 조건

t-test 의 전제조건인 **자료의 정규분포**를 만족하여야 한다.

**정규성 검정**은 t-test 하기 전에 반드시 시행하여야 한다.

예제 자료에 대한 정규성 검정을 시행한 결과 정규분포의 가정을 만족한다.



dBSTAT 에서 자동으로 정규성 검정을 시행한다.

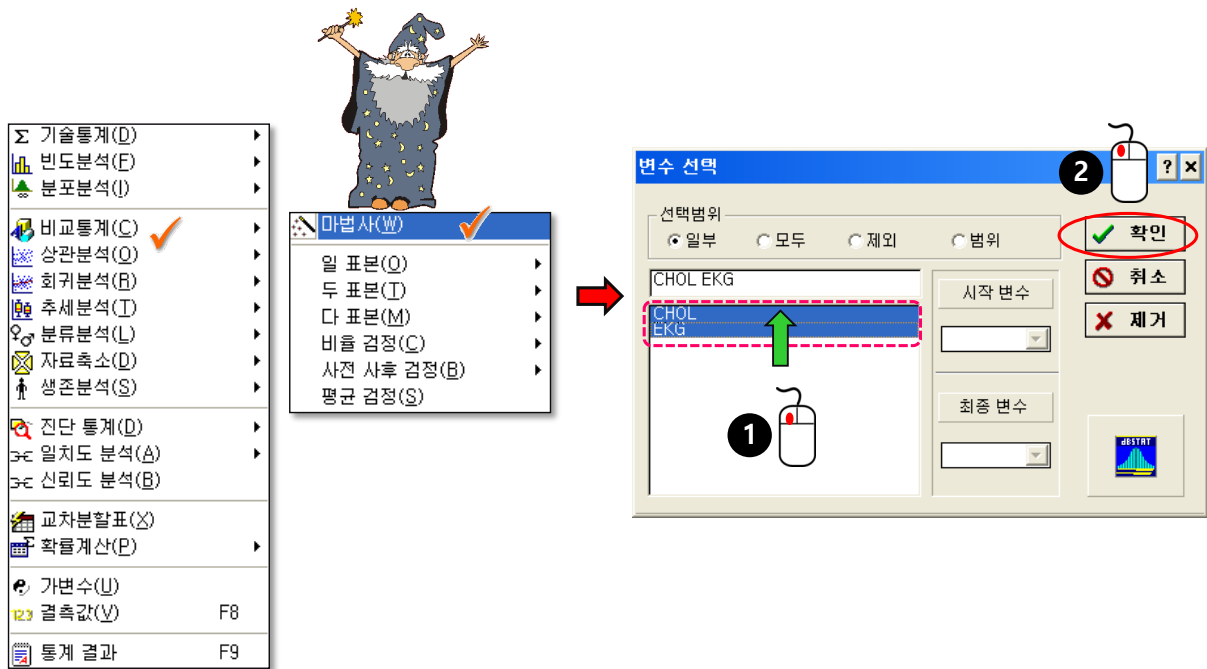


# dBSTAT

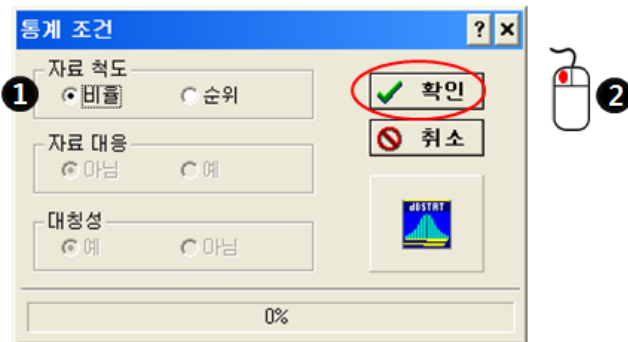
Data: two\_t.dbf

## 비교통계

1 메뉴 → [통계] ▶ [비교통계] ▶ [마법사] 메뉴를 선택한다.  
[변수 선택] 대화상자가 나타난다.



- ① [변수 선택] 대화상자에서 변수 CHOL, EKG 를 선택한다.
- ② [확인] 버튼을 클릭하여 선택 작업을 마치면 [통계 조건] 대화상자가 나타난다.



- ① [통계 조건] 대화상자에서 [자료 척도] ⊙[비율] 항목을 선택한다.
- ② [확인] 버튼을 클릭하여 선택 작업을 마치면 통계결과가 출력된다.

## 2 독립표본 t 검정법

dBSTAT에서는 [통계 조건]에 따라서 [마법사]가 적절한 통계검정을 시행한다.



[통계] ▶ [비교통계] ▶ [두 표본] ▶ [독립표본 t 검정]을 선택하여도 동일한 결과가 출력된다.

통계 결과

통계결과    통계치    그래프

🔍 검색    🖨️ 인쇄    💾 저장    🛑 종료

Independent t-test

변 수 명: CHOL 와 EKG

---

집 단 명	표본크기	평균값	표준편차	표준오차
1	10	166.40	29.35	9.28
2	10	239.20	37.92	11.99

(집단간 차이) 평균값 = -72.8000    신뢰구간(95%)    -104.6602 ~ -40.9398    3

1 [등분산성 검정]

F test	P = 0.2286	F = 1.67	( $\alpha = 0.05$ ): F = 4.03
Levene test	P = 0.4407	F = 0.62	( $\alpha = 0.05$ ): F = 5.98

2 [유의성 검정]

통 계 량:	Student t = 4.8005	자유도: df = 18
유의수준 ( $\alpha = 0.05$ ):	t = 2.1009	확 률: P = 0.0001

결 론: 비교 집단(평균값) 간에 통계적으로 유의한 차이가 있습니다.

4 [정규분포 검정]

집단변수: 1	Shapiro-Wilk W = 0.9002	확 률: P = 0.2100
결 론:	정규분포를 이룹니다.	

집단변수: 2	Shapiro-Wilk W = 0.9583	확 률: P = 0.7526
결 론:	정규분포를 이룹니다.	

① 등분산성 검정 F-test, Levene test 결과  $p > 0.05$  로 등분산을 만족한다.

② 유의성 검정

t-value = 4.8, 유의수준  $\alpha = 0.05$ , df = 18  $\Rightarrow$  임계값  $t_{0.05, 18} = 2.10$

양측검정 유의확률  $p = 0.0001$

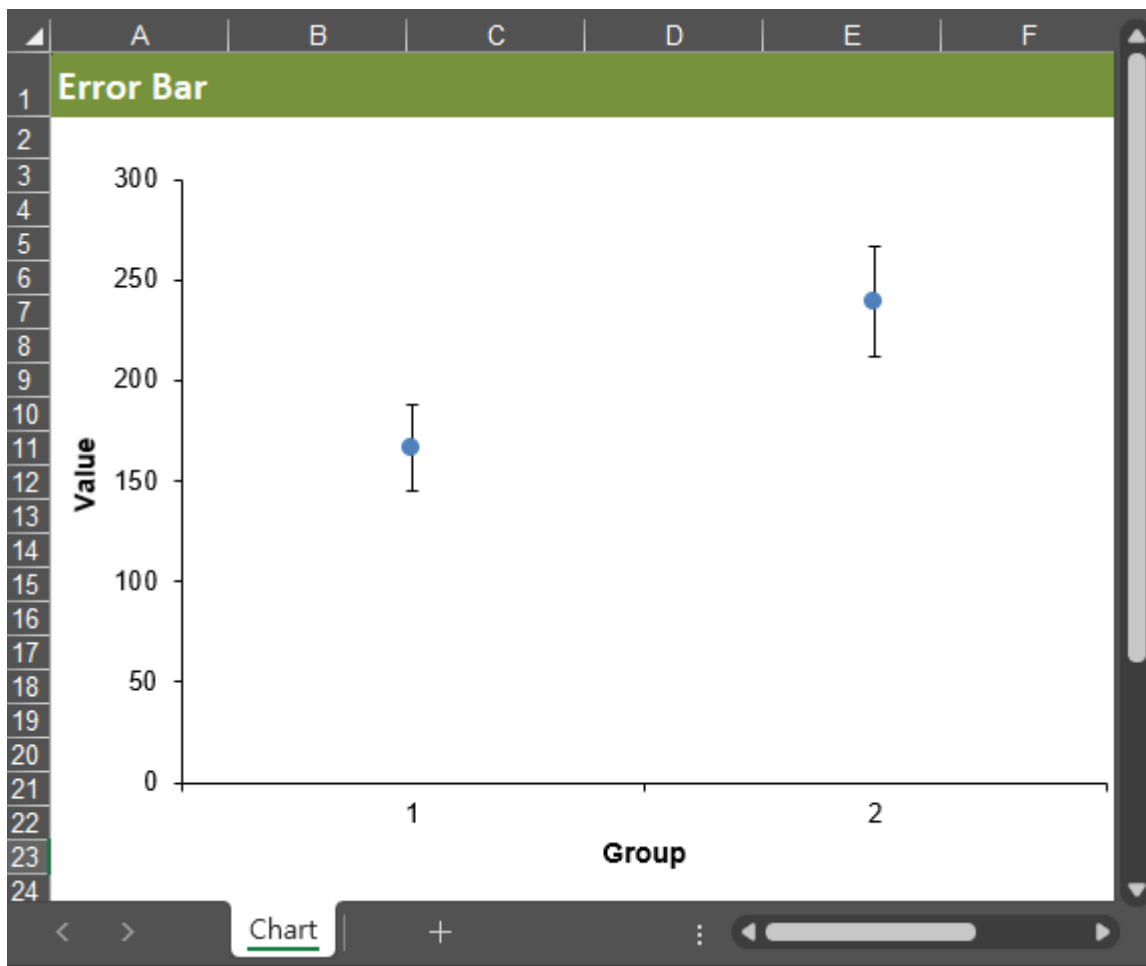
③ 95% 신뢰구간 = (하한값, 상한값) = (-104.7, -40.9)

④ 정규분포 검정

Shapiro-Wilk test 결과  $p > 0.05$  이므로 정규분포 가정을 만족한다.

### 3 Excel Chart

**Error Bar:** Excel 창에서 오차막대(mean with 95% 신뢰구간)이 나타난다.






## Exercise

### Paired $t$ -test



사과(apple) 다이어트를 12 명의 성인에게 시행하여 다이어트 하기 전과 1 개월후에 혈중 콜레스테롤(단위: mg/dL)을 측정하였다. 다이어트 전후에 평균 콜레스테롤은 차이가 있는가? 통계적 검정을 하시오.

	콜레스테롤 (mg/dL)		
	Before	After	Difference*
1	270	218	-52
2	236	234	-2
3	210	214	4
4	142	116	-26
5	280	200	-80
6	272	276	4
7	160	146	-14
8	220	182	-38
9	226	238	12
10	242	288	46
11	186	190	4
12	266	236	-30

\*  $After - Before$

#### [통계소프트웨어 사용]

##### 1. 통계적 검정법

① 표본평균 검정, ② 짝표본, ③ 표본수 = 2  $\Rightarrow$  **paired t-test** 를 선택한다.

##### 2. 전제 조건

① t-test 의 전제조건인 **자료의 정규분포**를 만족하여야 한다.

② **정규성 검정**은 t-test 하기 전에 반드시 시행하여야 한다.

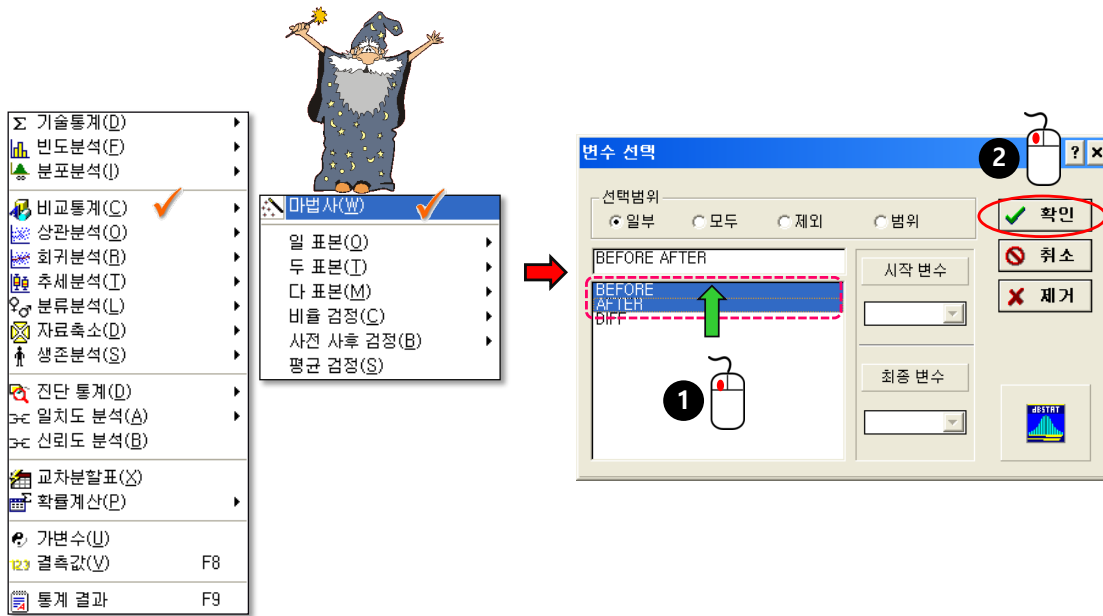
예제에서 **Difference** 에 대한 정규성 검정을 시행한 결과 정규분포의 가정을 만족한다.



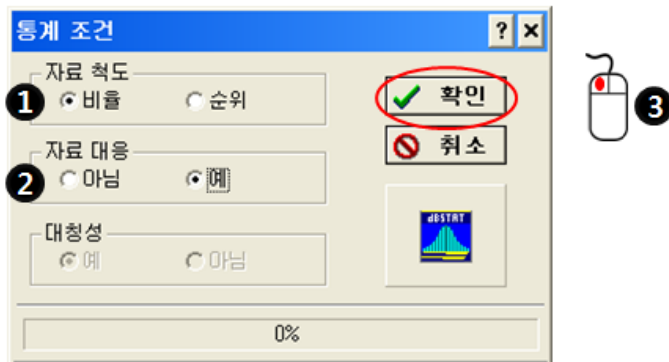
# dBSTAT

Data: paired\_t.dbf

1 메뉴 → [통계] ▶ [비교통계] ▶ [마법사] 메뉴를 선택한다.  
[변수 선택] 대화상자가 나타난다.



- ① [변수 선택] 대화상자에서 변수 BEFORE, AFTER 를 선택한다.
- ② [확인] 버튼을 클릭하여 선택 작업을 마치면 [통계 조건] 대화상자가 나타난다.



- ① [통계 조건] 대화상자에서 [자료 척도]  [비율] 항목을 선택한다.
- ② [통계 조건] 대화상자에서 [자료 대응]  [예] 항목을 선택한다
- ③ [확인] 버튼을 클릭하여 선택 작업을 마치면 통계결과가 출력된다.

## 2 짝표본 t 검정법

dBSTAT 에서는 [통계 조건]에 따라서 [마법사]가 적절한 통계검정을 시행한다.



[통계] ▶ [비교통계] ▶ [두 표본] ▶ [짝표본 t 검정]을 선택하여도 동일한 결과가 출력된다.

통계결과 | 통계치 | 그래프

🔍 검색 | 🌐 인쇄 | 💾 저장 | 🛑 종료

Paired t-test

변 수 명: BEFORE 와 AFTER

---

집 단 명	표본크기	평균값	표준편차	표준오차
BEFORE	12	225.83	44.95	12.98
AFTER	12	211.50	49.31	14.23

(집단간 차이) 평균값 = 14.3333      신뢰구간(95%): -6.7976 ~ 35.4643 ②

① [유의성 검정]

통 계 량: t = 1.4930      자유도: df = 11  
 유의수준 (α = 0.05): t = 2.2010      확 률: P = 0.1636

결 론: 비교 집단(평균값) 간에 통계적으로 유의한 차이가 없습니다.

③ [정규분포 검정]

변 수: BEFORE - AFTER  
 Shapiro-Wilk W = 0.9688      확 률: P = 0.8461  
 결 론: 정규분포를 이룹니다.

### ① [유의성 검정]

t-value = 1.49, 유의수준  $\alpha = 0.05$ , df = 11  $\Rightarrow$  임계값  $t_{0.05, 11} = 2.20$

양측검정 유의확률  $p = 0.16$

### ② 95% 신뢰구간 = (하한값, 상한값) = (-6.8, 35.5)

### ③ [정규분포 검정]

Shapiro-Wilk test 결과  $p > 0.05$  이므로 정규분포 가정을 만족한다.

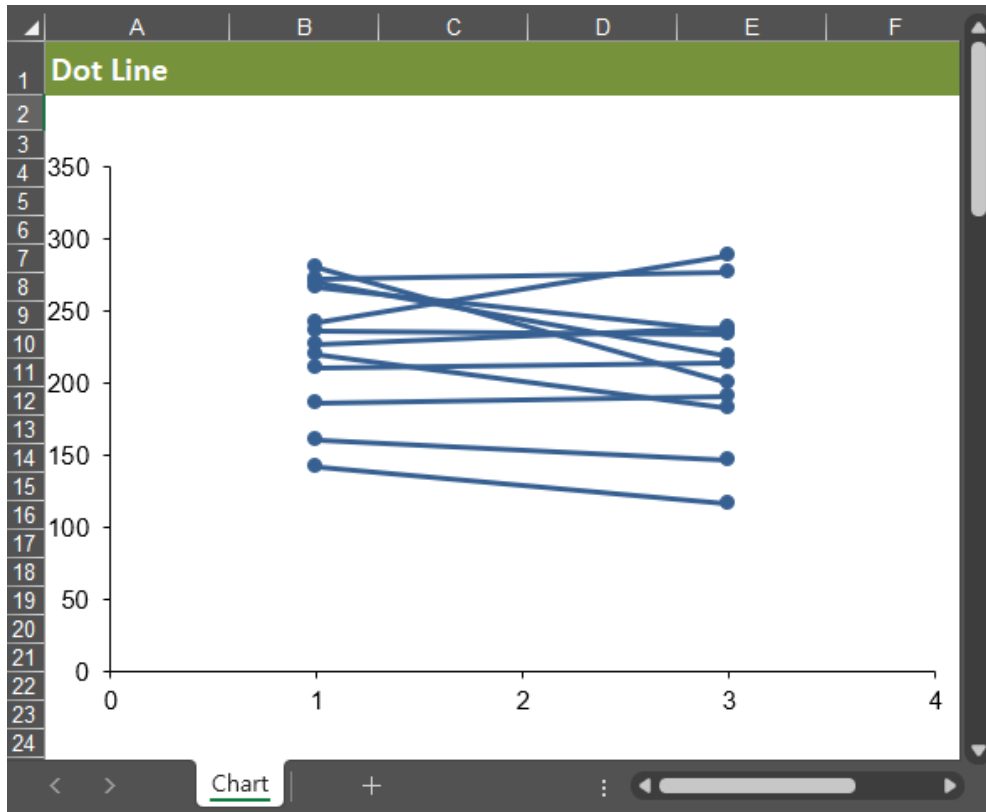
[해석]

✎ t-value 1.49 < t 임계값 (2.20) 이므로  $P > 0.05$  로 통계적으로 유의한 차이가 없다고 판정한다.

✎  $p = 0.164$  이므로 통계적으로 유의한 차이가 없다고 판정한다.

### 3 Excel Chart

**Dot Line:** Excel 창에서 Dot line plot 이 나타난다.



짝을 이룬 두 변수(BEFORE, AFTER)의 자료를 점으로 표시한 다음 짝끼리 서로 연결하여 그래프로 그린다.




## Exercise

### Mann-Whitney Test



산모의 흡연이 저체중 신생아의 건강상태에 미치는 영향을 알아보기 위하여 흡연자와 비흡연자 각각 10 명에서 저체중아 출산 1 분후 아프가 점수(Apgar score)를 측정하였다. 비흡연 산모의 신생아 아프가 점수가 흡연 산모의 신생아에 비하여 차이가 있는가? 유의수준 5%에서 통계적 검정을 하시오.

Smoking	Apgar Score									
 비흡연 산모	5	3	0	4	1	2	6	10	3	7
 흡연 산모	5	7	1	0	3	1	0	4	2	3

#### [해설]

##### 1) 가설 설정

- ① 귀무가설  $H_0$ :  $Median_1 = Median_2$  (두 표본의 모집단 분포(중앙값)가 동일하다.)
- ② 대립가설  $H_1$ :  $Median_1 \neq Median_2$  (두 표본의 모집단 분포(중앙값)가 다르다.)

##### 2) 통계적 검정법



Apgar 점수는 순서자료이므로 독립 두 표본 비교에 사용되는 **Mann-Whitney test** 를 사용한다.



# dBSTAT

Data: Mann-Whitney.dbf

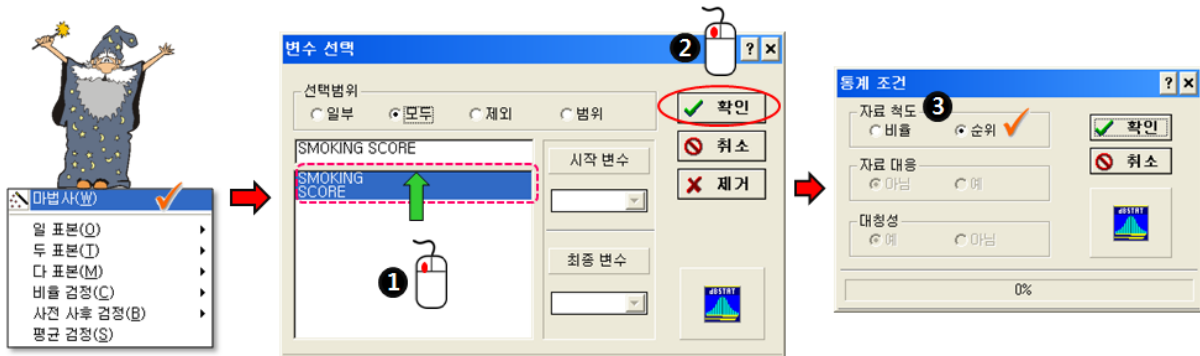
1 메뉴 → [통계] ▶ [비교통계] ▶ [마법사] 메뉴를 선택한다.



[통계] ▶ [비교통계] ▶ [두 표본] ▶ [맨-휘트니 검정]을 선택하여도 동일한 결과가 출력된다.



자료에 0 이 있으므로 [통계] ▶ [결측값] ▶ [결측값 설정] 결측값 0 → -9 로 바꾸어야 한다.



- ① [변수 선택] 대화상자에서 변수 SMOKING, SCORE 를 선택한다.
- ② [확인] 버튼을 클릭하여 선택 작업을 마치면 [통계 조건] 대화상자가 나타난다.
- ③ [통계 조건] 대화상자에서 [자료 척도] [순위] 항목을 선택한다. [확인] 버튼을 누른다.

## 2 통계결과

통계결과    통계치    그래프

🔍 검색    📄 인쇄    💾 저장    🛑 종료

Mann-Whitney U (Wilcoxon Rank Sum) Test

---

자료수  
 10 SMOKING = 1  
 10 SMOKING = 2  
 ----  
 20 합 계

1 [대표본 근사검정]  
 Mann-Whitney U = 35.00                      Wilcoxon W = 90.0000

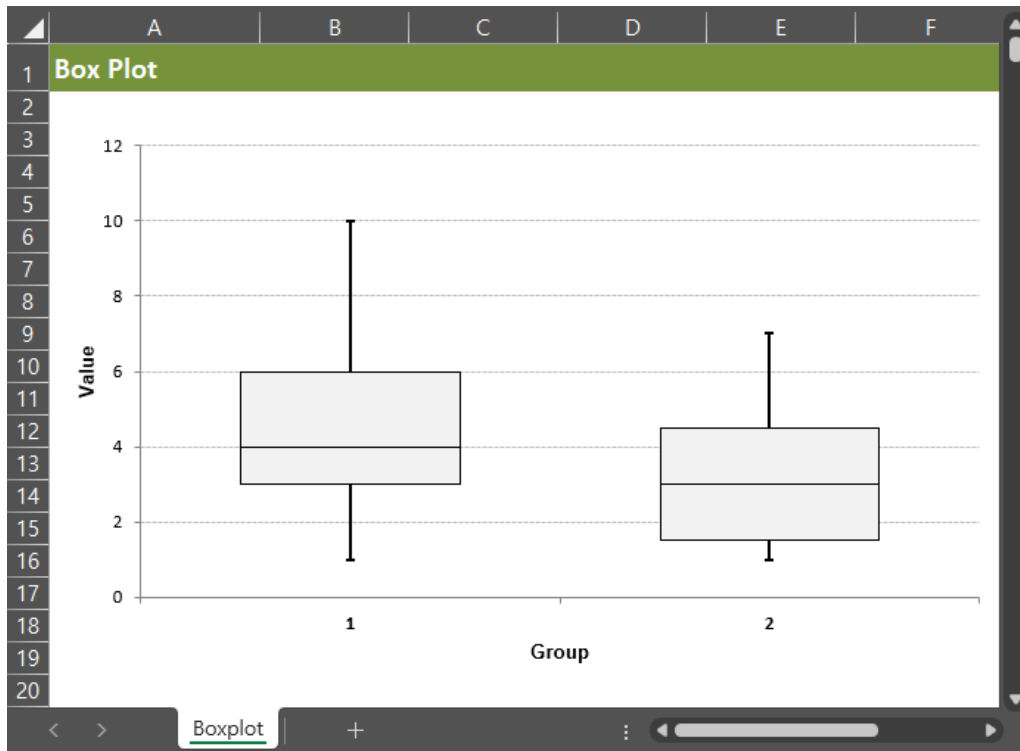
2 통계 량: Z = 1.1339                      동순위 교정후: Z = -1.1434  
 유의수준 (α = 0.05): Z = 1.96                      확 률: P = 0.2529

결 론: 비교 집단 분포(중앙값) 간에 통계적으로 유의한 차이가 없습니다.

- ① 대표본 근사검정 Mann-Whitney U = 35, Wilcoxon W = 90
- ② 통계량  $Z = -1.1434$ ,  $p = 0.2529$  ( $p > 0.05$ )로 유의한 차이가 없다고 결론을 내린다.

### 3 Excel Chart

**Box Plot:** Excel 창에서 Box plot 이 나타난다.






## Exercise

### Wilcoxon Signed Rank Test



음악요법이 심리상태에 미치는 영향을 알아보기 위하여 7명에서 음악요법 시행 전후에 행복지수를 측정하였다. 음악요법이 효과가 있는지 검정하시오.

	Score		
	Before	After	Difference*
1	3	7	4
2	2	8	6
3	5	4	-1
4	6	9	3
5	5	10	5
6	1	9	8
7	8	9	1

\* *After - Before*



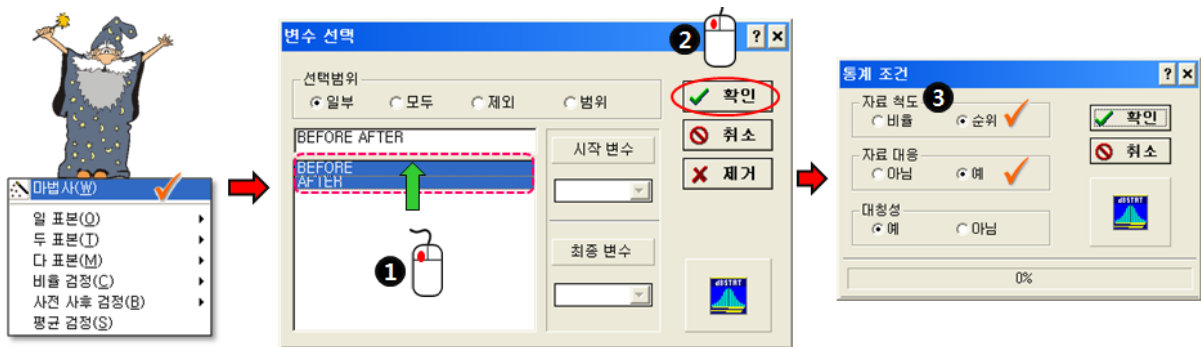
dBSTAT

Data: Wilcoxon.dbf

1 메뉴 → [통계] ▶ [비교통계] ▶ [마법사] 메뉴를 선택한다.



[통계] ▶ [비교통계] ▶ [두 표본] ▶ [윌콕슨 부호순위검정]을 선택하여도 동일하다.

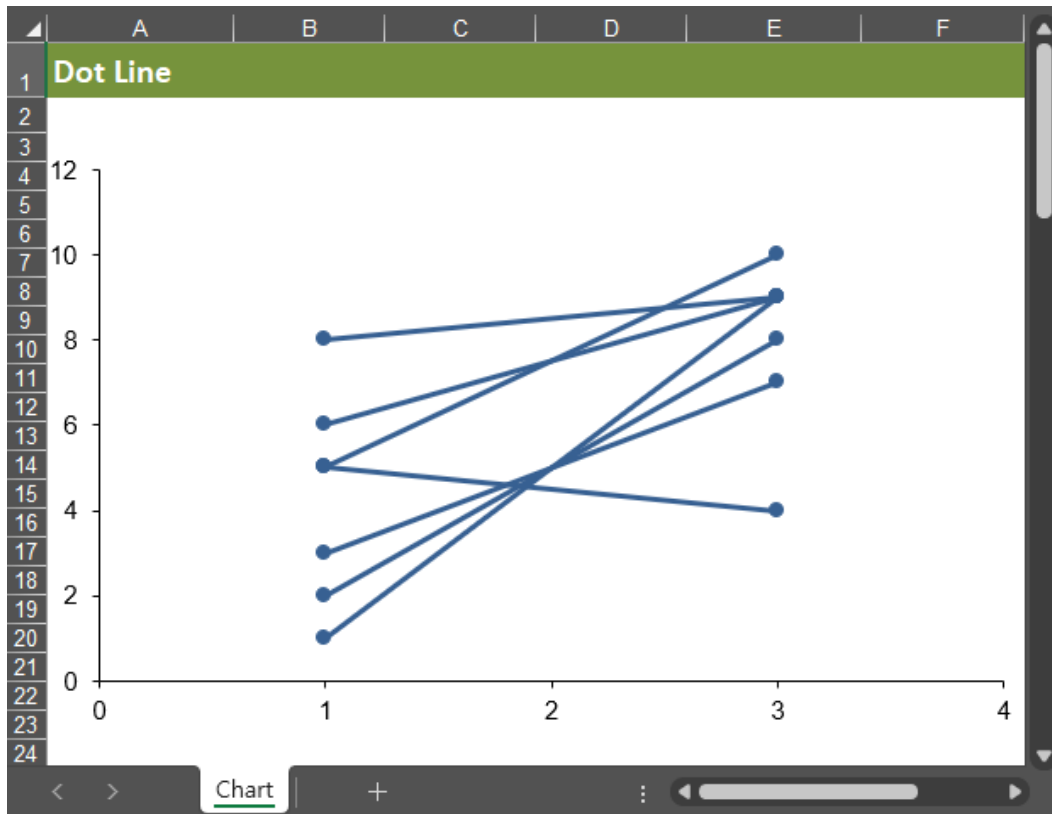


- ① [변수 선택] 대화상자에서 변수 BEFORE, AFTER 를 선택한다.
- ② [확인] 버튼을 클릭하여 선택 작업을 마치면 [통계 조건] 대화상자가 나타난다.
- ③ [통계 조건] 대화상자에서 [자료 척도] [순위], [자료 대응] [예] 항목을 선택한다



### 3 Excel Chart

**Dot Line:** Excel 창에서 Dot line plot 이 나타난다.



# Multi-Sample





## Exercise

### ANOVA Test



식생활(Diet) 습관에 따라 1 = 채식 2 = 균형식 3 = 육식 3 가지로 나누어 각각 10 명에서 혈중 콜레스테롤(단위: mg/dL)을 측정한 자료이다. 세 집단의 평균 콜레스테롤은 차이가 있는가? 유의수준 5%에서 검정하시오.

Cholesterol					
Diet 1		Diet 2		Diet 3	
141	136	202	213	206	241
189	186	256	237	210	258
173	177	183	249	226	270
198	124	194	220	250	293
204	136	184	271	255	328

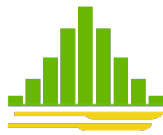
#### [해설]

##### 1. 통계적 검정법

- ① 표본평균 검정
- ② 독립표본
- ③ 집단수 = 3  $\Rightarrow$  **one-way ANOVA** 를 선택한다.

##### 2. 전제 조건

- ① 정규성 검정
- ② 등분산 검정



# dBSTAT

Data : anova.dbf

1 메뉴 → [통계] ▶ [비교통계] ▶ [마법사] 메뉴를 선택한다. [변수 선택] 대화상자가 나타난다.



[통계] ▶ [비교통계] ▶ [다 표본] ▶ [일원분산분석]을 선택하여도 동일한 결과가 출력된다.

The screenshot shows the '통계(S)' menu with '비교통계(C)' selected. A sub-menu '마법사(W)' is open, showing '일 표본(O)', '두 표본(I)', '다 표본(M)', '비율 검정(C)', '사건 사후 검정(B)', and '평균 검정(S)'. A red arrow points to the '변수 선택' dialog box. In this dialog, '선택범위' is set to '일부', and 'CHOL' and 'DIET' are selected in the variable list. The '확인' button is circled in red, and a mouse cursor is shown clicking it. A '1' is placed over the variable list, and a '2' is placed over the '확인' button.

① [변수 선택] 대화상자에서 변수 CHOL, DIET 를 선택한다.

② [확인] 버튼을 클릭하여 선택 작업을 마치면 [통계 조건] 대화상자가 나타난다.

The screenshot shows the '통계 조건' dialog box. Under '자료 척도', the '비율' radio button is selected. The '확인' button is circled in red, and a mouse cursor is shown clicking it. A '1' is placed over the '비율' radio button, and a '2' is placed over the '확인' button.

① [통계 조건] 대화상자에서 [자료 척도] Ⓞ[비율] 항목을 선택한다.

② [확인] 버튼을 클릭하여 선택 작업을 마치면 통계결과가 출력된다.

## 2 일원분산분석

통계결과    통계치    그래프

🔍 검색    📄 인쇄    💾 저장    🛑 종료

Analysis of Variance (ANOVA) test

CHOL 와 DIET ( 1 부터 3 )

-----

1 일원분산분석

변인	자유도	제공합	분산	분산비
집단간 (처리)	2	38891.27	19445.63	18.1262
집단내 (오차)	27	28965.40	1072.79	
합계	29	67856.67		

2 유의수준 ( $\alpha = 0.05$ ):  $F = 3.35$                       확률:  $P = 0.0000$

3 결론: 집단들 간의 평균값이 동일하지 않습니다.

① 집단간:집단내 분산비  $F = 18.13$

② 양측검정 유의확률  $p = 0.0000$  ( $p < 0.0001$ )

③  $p < 0.05$  이므로 집단들간 평균이 통계적으로 유의한 차이가 있다고 판정한다.

### 3 사후검정

**1** [Scheffe 다중비교]

집 단 명	1	2	3
1			
2	0.0037* (-54.50, -92.41~-16.59)		
3	0.0000* (-87.30, -125.21~-49.39)	0.1003 (-32.80, -70.71~5.11)	

\* 표시: 검정확률  $P < 0.05$   
 () 표시: 평균차, 95% 신뢰구간

**2** [Bonferroni 다중비교]

집 단 명	1	2	3
1			
2	0.0027* (-54.50, -91.90~-17.10)		
3	0.0000* (-87.30, -124.70~-49.90)	0.1008 (-32.80, -70.20~4.60)	

\* 표시: 검정확률  $P < 0.05$   
 () 표시: 평균차, 95% 신뢰구간

- ① Scheffe 검정 쌍별 비교에서 유의확률 p와 평균차, 평균차에 대한 95% 신뢰구간을 보여준다.
- ② Bonferroni 검정에서 유의확률 p와 함께 평균차, 평균차에 대한 95% 신뢰구간을 보여준다.

Diet 1 과 Diet 2, Diet 1 과 Diet 3 집단별 비교에서  $p < 0.05$  이므로 유의한 차이가 있다.

## 4 분산분석 가정

The screenshot shows the SPSS '통계결과' (Statistical Results) window. It has tabs for '통계결과', '통계치', and '그래프'. Below the tabs are buttons for '검색' (Search), '인쇄' (Print), '저장' (Save), and '종료' (End). The main content area is divided into two sections:

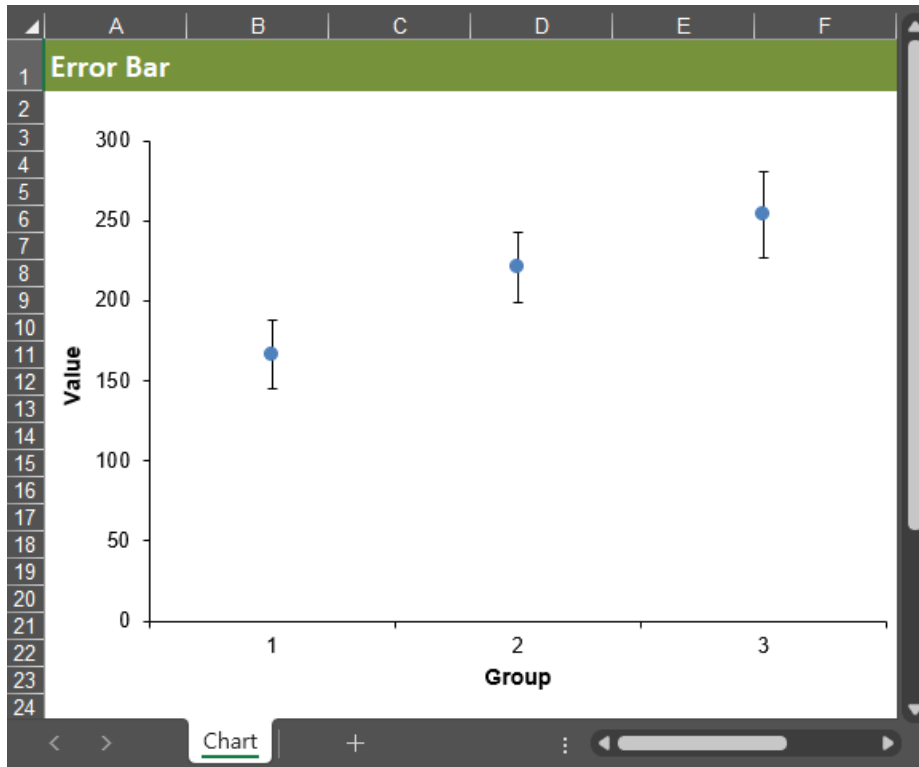
**1 [등분산성 검정]**  
Bartlett test P = 0.7628 (α = 0.05):  $\chi^2 = 5.991$   
결론: 집단 간의 분산이 동일합니다.  
Levene test P = 0.9817 F = 0.02  
결론: 집단 간의 분산이 동일합니다.

**2 [정규분포 검정]**  
집단변수: 1  
Shapiro-Wilk W = 0.9002 확률: P = 0.2100  
결론: 정규분포를 이룹니다.  
집단변수: 2  
Shapiro-Wilk W = 0.9429 확률: P = 0.5695  
결론: 정규분포를 이룹니다.  
집단변수: 3  
Shapiro-Wilk W = 0.9528 확률: P = 0.6864  
결론: 정규분포를 이룹니다.

- ① 등분산성 검정: Bartlett test, Levene test 결과  $p > 0.05$  로 등분산을 만족한다.
- ② 정규분포 검정: Shapiro-Wilk test 결과  $p > 0.05$  이므로 정규분포 가정을 만족한다.

## 5 Excel Chart

**Error Bar:** Excel 창에서 오차막대(mean with 95% 신뢰구간)이 나타난다.



- ① 일원분산분석은 평균과 95% 신뢰구간으로 표현되는 오차막대가 적합한 그래프이다.
- ② X 축은 Diet 1, 2, 3 세 집단을 나타내며 Y 축은 콜레스테롤 평균과 95% 신뢰구간을 보여준다.







## Exercise

## Two-way ANOVA Test



식생활과 운동이 혈압에 미치는 효과를 보기 위하여 고혈압 환자들을 대상으로 Diet 방법(1=채식, 2=즉석식품)과 2 가지 운동(1=약함, 2=강함)을 5 개월 시행한 후에 수축기 혈압을 측정하였다. 두 요인에 따라 혈압의 차이가 있는지 검정하시오.

Factor levels		Factor B (Diet)									
		1 					2 				
Factor A (Exercise)	1 	130 142 131 124 124	157 143 162 150 149								
		131 143 131 135 140	140 159 158 160 155								
	2 	120 131 122 129 120	132 128 142 131 135								
		119 134 123 125 130	120 139 133 135 140								



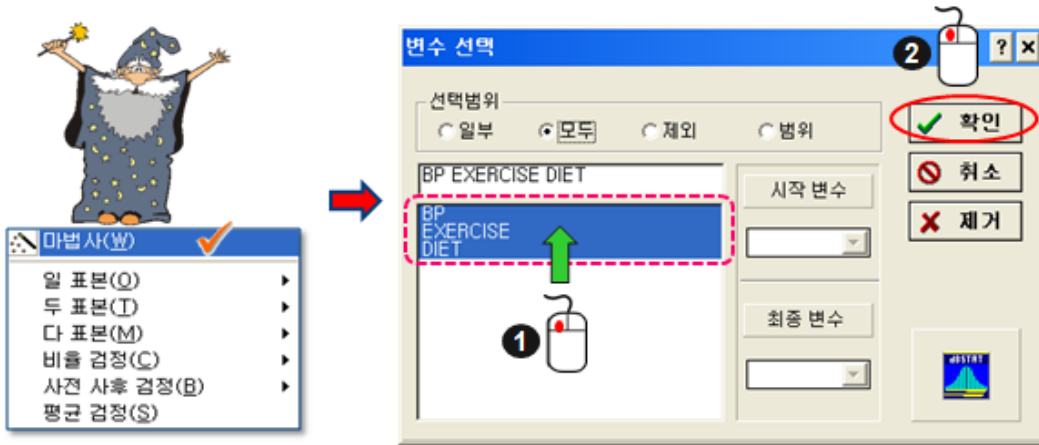
# dBSTAT

Data : anova2.dbf

1 메뉴 → [통계] ▶ [비교통계] ▶ [마법사] 메뉴를 선택한다. [변수 선택] 대화상자가 나타난다.



[통계] ▶ [비교통계] ▶ [다 표본] ▶ [이원분산분석]을 선택하여도 동일한 결과가 출력된다.



① [변수 선택] 대화상자에서 변수 BP, EXERCISE, DIET 를 선택한다.

② [확인] 버튼을 클릭하여 선택 작업을 마치면 통계결과가 출력된다.

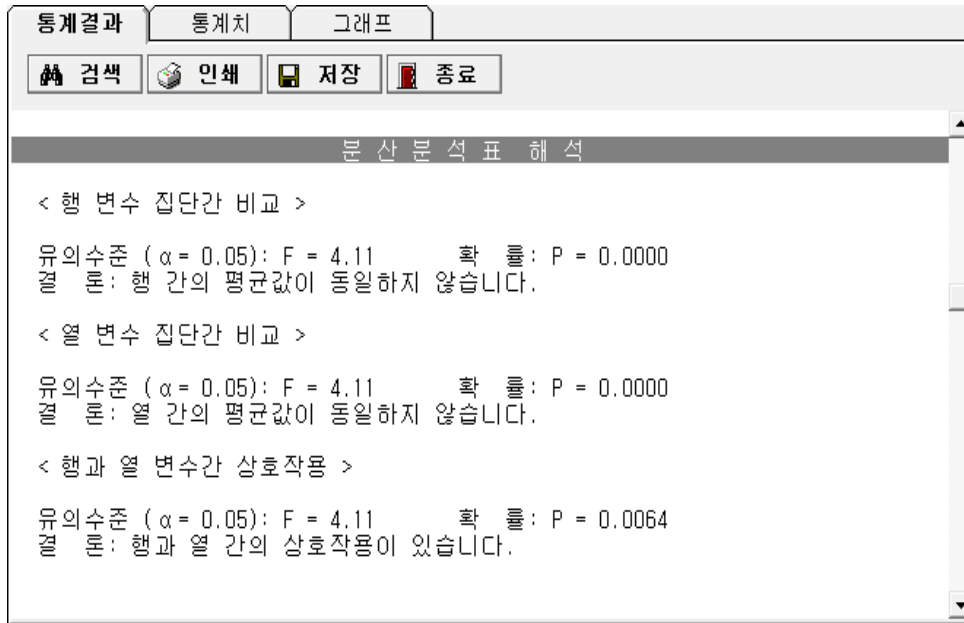
## 2 이원분산분석

통계결과		통계치	그래프
🔍	검색	📄	인쇄
💾	저장	🔴	종료
Analysis of Variance (ANOVA) test			
변 수 명 : 행 - EXERCISE 열 - DIET			
이원 분산 분석 (요인설계)			
변동요인	자유도	제 곱 합	분 산
① 행 간	1	1904.40	1904.40
② 열 간	1	2016.40	2016.40
③ 상호작용	1	360.00	360.00
처 리	3	4280.80	1426.93
오 차	36	1547.60	42.99
합 계	39	5828.40	

① 행간 요인: Exercise 효과 분산비 F = 44.30

- ② 열간 요인: Diet 효과 분산비  $F = 46.91$
- ③ 상호작용: Exercise 와 Diet 상호작용 효과 분산비  $F = 8.37$

### 3 분산분석표 해석

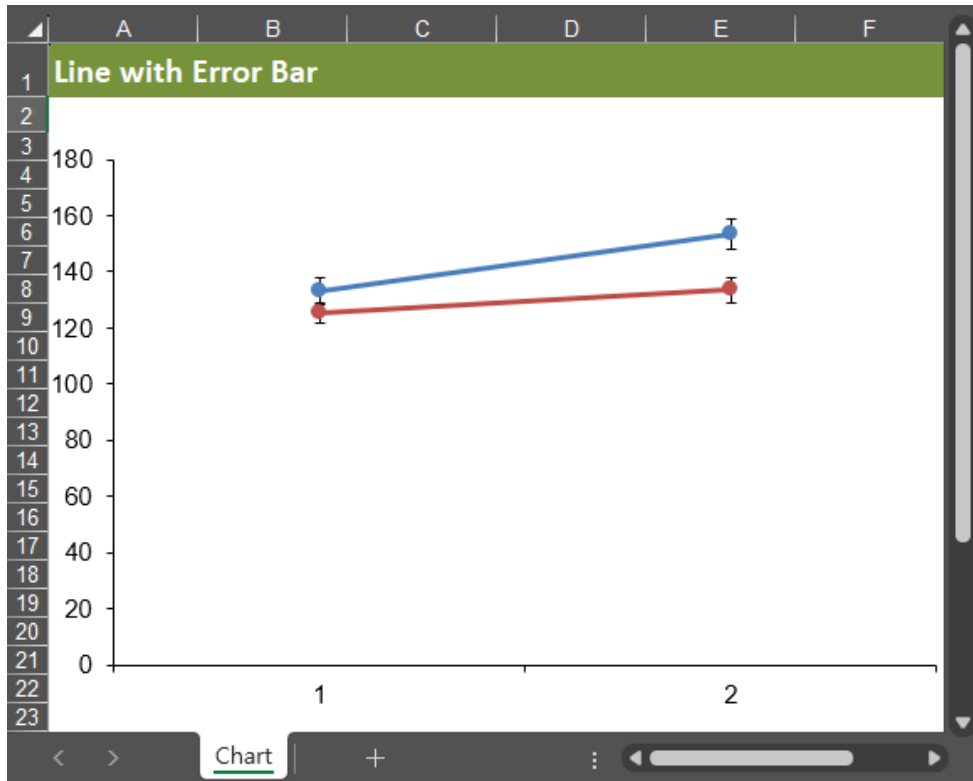


위의 분산분석표(ANOVA table)에 대한 통계적 해석을 내려준다.

- ① 행간 요인 (Exercise);  $p = 0.0000$  이므로 주효과가 있다.
- ② 열간 요인 (Diet);  $p = 0.0000$  이므로 주효과가 있다.
- ③ 상호작용 (Exercise\*Diet);  $p = 0.0064$  이므로 상호작용 효과가 있다.

## 4 Excel Chart

**Line with Error Bar:** Excel 창에서 오차막대가 선으로 연결된 plot 이 나타난다





## Exercise

### Repeated Measures ANOVA



비만 환자 10 명에서 Diet 시행 전, 시행 후 3 개월, 시행 후 6 개월에 혈중 콜레스테롤 농도를 측정하였다. 평균 콜레스테롤에 차이가 있는지 유의수준 5%에서 검정하시오. 전제조건인 정규성과 등분산성 검정은 생략하기로 한다.

Cholesterol		
Before Diet	3 M After	6 M After
206	202	141
210	256	189
226	183	173
250	194	198
255	184	204
241	213	136
258	237	186
270	249	177
293	220	124
328	271	136

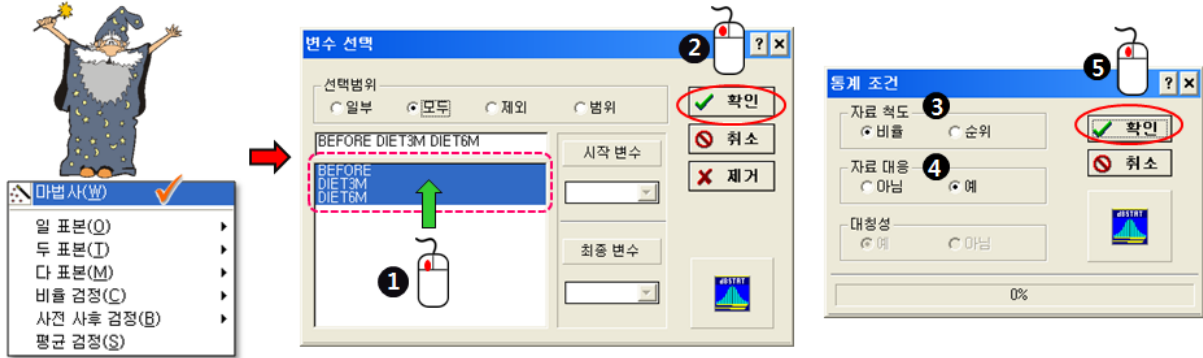


# dBSTAT

Data: ranova.dbf

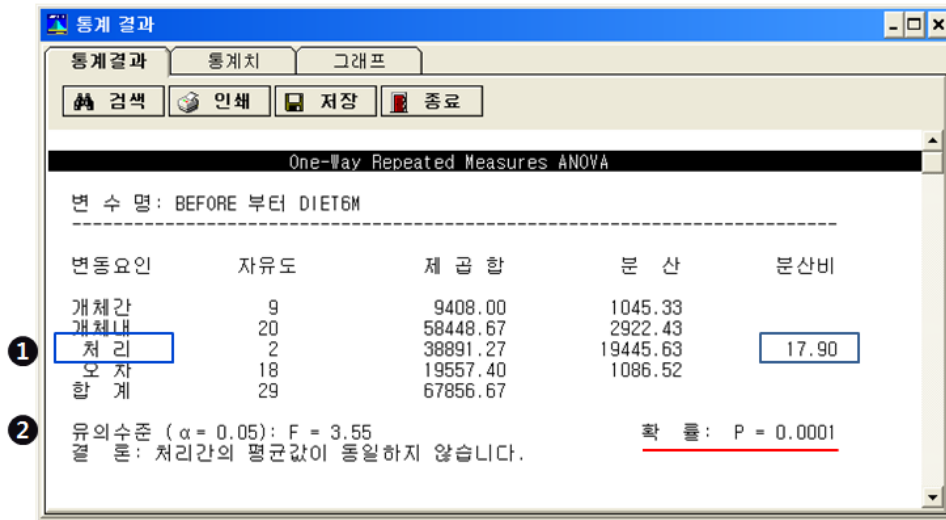
1 메뉴 → [통계] ▶ [비교통계] ▶ [마법사] 메뉴를 선택한다. [변수 선택] 대화상자가 나타난다.

💡 [통계] ▶ [비교통계] ▶ [다 표본] ▶ [반복측정 분산분석]을 선택하여도 동일하다.



- ① [변수 선택] 대화상자에서 변수 BEFORE, DIET3M, DIET6M 를 선택한다.
- ② [확인] 버튼을 클릭하여 선택 작업을 마치면 [통계조건] 대화상자가 나타난다.
- ③ [자료척도] [비율] 항목을 선택한다.
- ④ [자료대응] [예] 항목을 선택한다.
- ⑤ [확인] 버튼을 클릭하여 선택 작업을 마치면 [통계결과] 창이 나타난다.

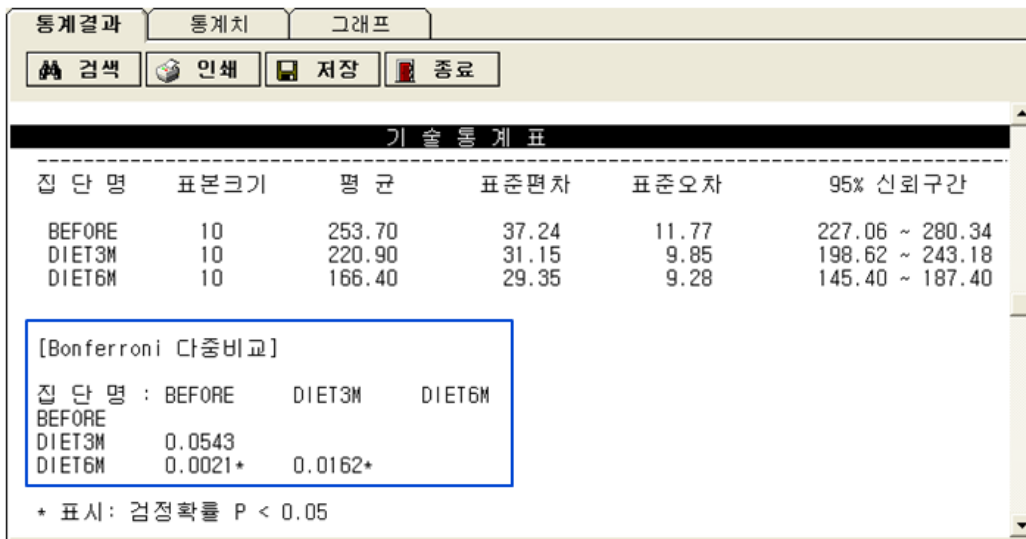
## 2 반복측정 분산분석



① 처리 요인 (Treatment): 분산비  $F = 17.90 \Rightarrow p = 0.0001$  이다.

②  $P < 0.05$  이므로 처리(Diet)에 따라 콜레스테롤 평균의 차이가 있다고 결론을 내린다.

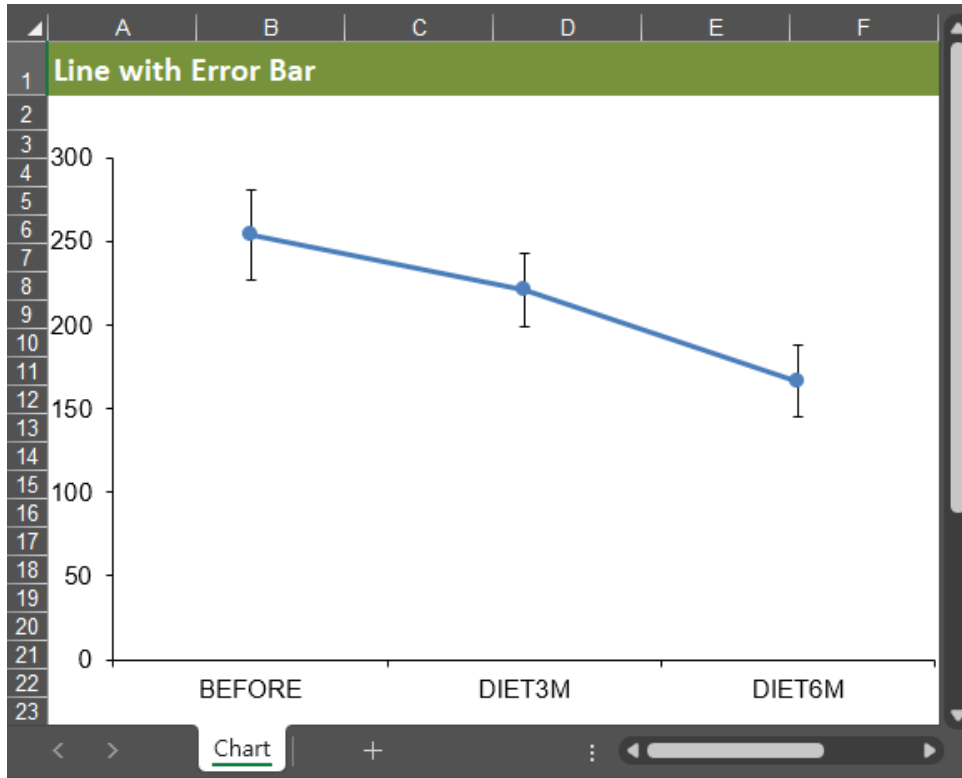
## 3 사후검정



Bonferroni 다중비교에서 Diet 전(Before)과 6개월 후, Diet 3개월 후와 6개월 후 집단간에 통계적으로 유의한 차이가 있다.

## 4 Excel Chart

**Line with Error Bar:** Excel 창에서 오차막대가 선으로 연결된 plot 이 나타난다







## Exercise

### ANCOVA Test



운동이 혈압(BP)에 미치는 효과를 보기 위하여 고혈압 환자들을 대상으로 2 가지 운동(1=약함, 2=강함)을 5 개월 시행한 후에 수축기 혈압을 측정하였다. 연령(Age)이 혈압에 미치는 효과를 조정하면서 운동 집단 간에 혈압의 차이를 검정하시오.

Exercise			
1 		2 	
BP	Age	BP	Age
158	41	144	39
185	60	138	45
152	41	145	47
159	47	162	65
176	66	142	46
156	47	170	67
184	68	124	42
138	43	158	67
172	68	154	56
168	57	162	64

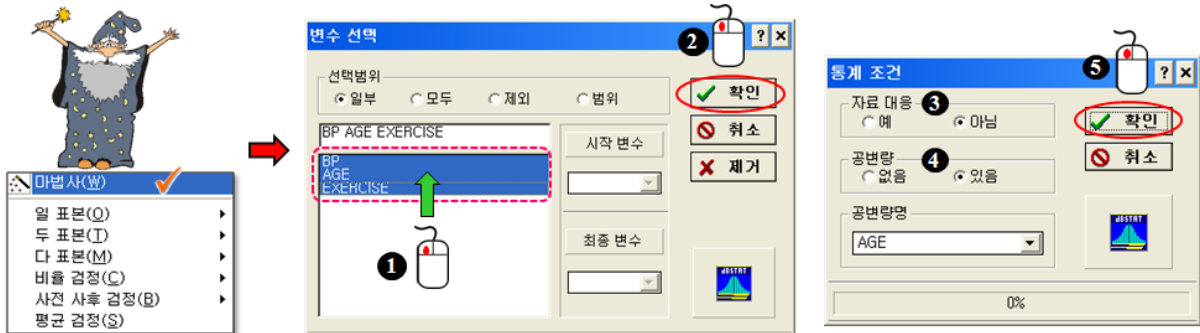


# dBSTAT

Data : ancova.dbf

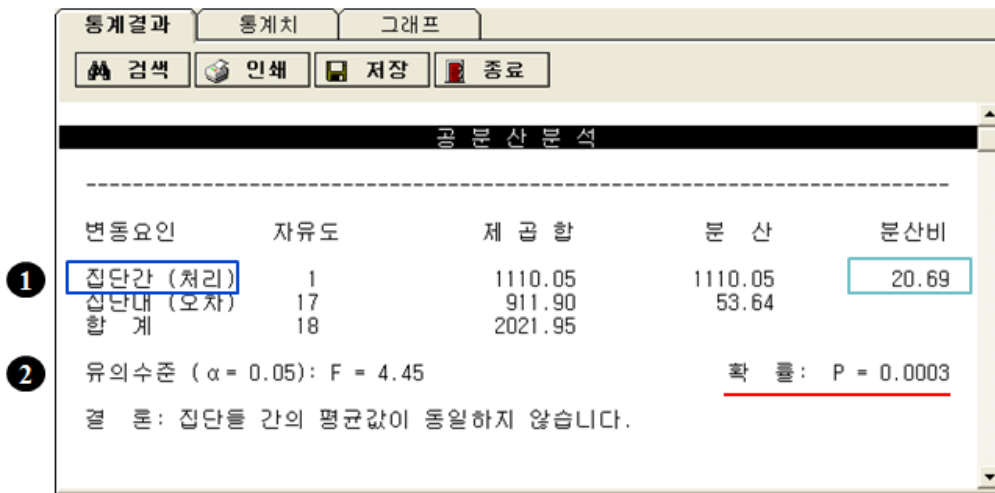
1 메뉴 → [통계] ▶ [비교통계] ▶ [마법사] 메뉴를 선택한다. [변수 선택] 대화상자가 나타난다.

💡 [통계] ▶ [비교통계] ▶ [다 표본] ▶ [공분산분석]을 선택하여도 동일한 결과가 출력된다.



- ① [변수 선택] 대화상자에서 변수 BP, AGE, EXERCISE 를 선택한다.
- ② [확인] 버튼을 클릭하여 선택 작업을 마치면 [통계 조건] 대화상자가 나타난다.
- ③ [통계 조건] 대화상자에서 자료대응  아님,
- ④ [공변수]  있음 ➔ [공변수명] AGE 선택한다.
- ⑤ [확인] 버튼을 클릭하여 선택작업을 마친다.

## 2 통계결과

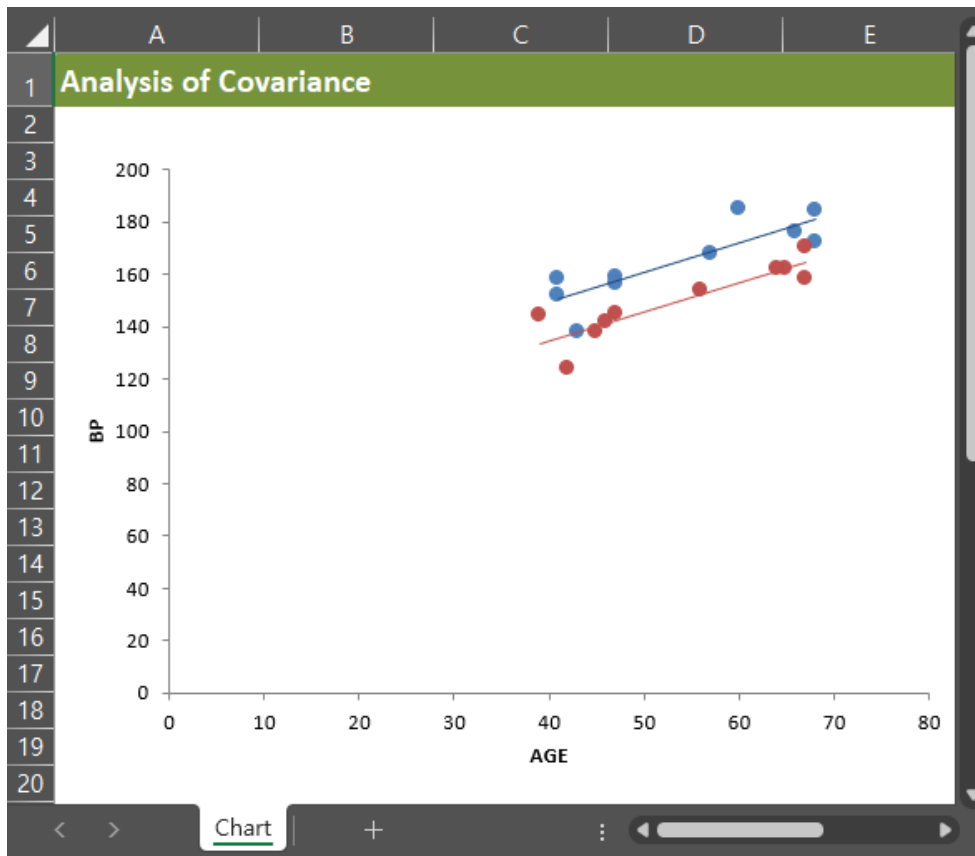


① 집단간 요인 (Exercise): 분산비  $F = 20.69$  이다.

②  $P = 0.003$  이므로 Exercise 집단 간 평균 혈압이 통계적으로 유의한 차이가 있다.

## 3 Excel Chart

**Scatter Plot:** Excel 창에서 EXERCISE group 별 scatter plot 이 나타난다.








## Exercise

### Kruskal-Wallis Test



식생활이 체중에 미치는 효과를 보기 위하여 Diet 방법에 따라 3 개의 집단으로 나누어 5 개월간 식이요법을 시행한 후에 체중감량(%)을 측정하였다. Diet 방법에 따라 체중감량의 효과에 차이가 있는지 유의수준 5%에서 검정하시오.

	Diet		
(Group) Independent	1 	2 	3 
Dependent variable (Weight)	25	18	12
	24	4	14
	22	6	7
	16	10	3
	20	8	5

Diet 1 = 채식, 2 = 균형식, 3 = 육식



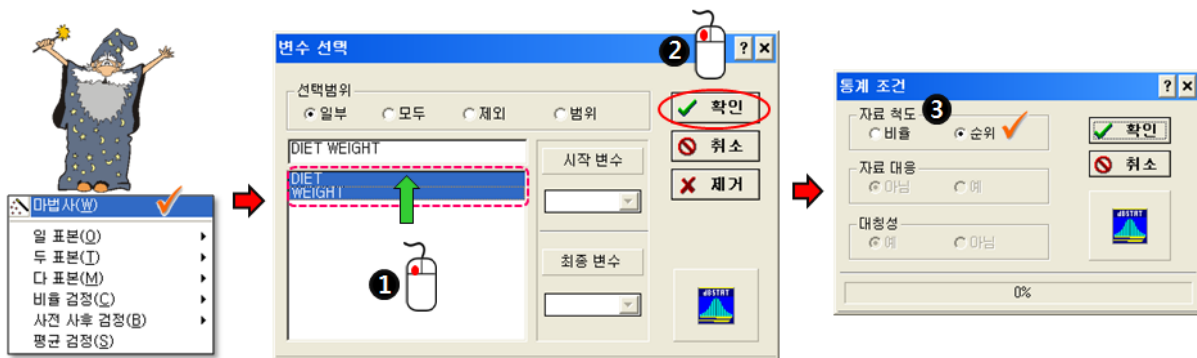
# dBSTAT

Data : Kruskal-Wallis.dbf

## 비교통계

1 메뉴 → [통계] ▶ [비교통계] ▶ [마법사] 메뉴를 선택한다.

[통계] ▶ [비교통계] ▶ [다 표본] ▶ [크루스칼-왈리스 검정]을 선택해도 동일한 결과가 출력된다.



- ① [변수 선택] 대화상자에서 변수 DIET, WEIGHT 를 선택한다.
- ② [확인] 버튼을 클릭하여 선택 작업을 마치면 [통계 조건] 대화상자가 나타난다.
- ③ [통계 조건] 대화상자에서 [자료 척도] [순위] 항목을 선택한다. [확인] 버튼을 누른다.

## 2 통계결과

통계결과 통계치 그래프

검색 인쇄 저장 종료

Kruskal-Wallis Test

자료수  
5 DIET = 1  
5 DIET = 2  
5 DIET = 3  
----  
15 합계

1 [소표본 정밀검정]  
통계량: Kruskal-Wallis H = 8.72  
유의수준 ( $\alpha = 0.05$ ): Kruskal-Wallis H = 5.78 확률:  $P < 0.05$   
결론: 비교 집단 분포(중앙값) 간에 통계적으로 유의한 차이가 있습니다.

2 [대표본 근사검정]  
통계량:  $\chi^2 = 8.72$  동순위 교정후:  $\chi^2 = 8.72$   
유의수준 ( $\alpha = 0.05$ ):  $\chi^2 = 5.99$  확률:  $P = 0.0128$   
결론: 비교 집단 분포(중앙값) 간에 통계적으로 유의한 차이가 있습니다.

① 소표본 정밀검정 Kruskal-Wallis H = 8.72 > 임계값 K-W H = 5.78  $\Rightarrow p < 0.05$  이다.

② 대표본 근사검정  $\chi^2 = 8.72 \Rightarrow p = 0.0128 < 0.05$

✎ 양측검정 확률  $p < 0.05$  이므로 "Diet 3 group 간에 체중감량은 동일하지 않다 (유의한 차이가 있다)."는 결론을 내린다.

## 3 사후검정

dBSTAT 에서 K-W 검정 결과가 통계적으로 유의하면 자동적으로 사후검정 결과를 보여준다.

통계결과 통계치 그래프

검색 인쇄 저장 종료

Dunn 다중비교표

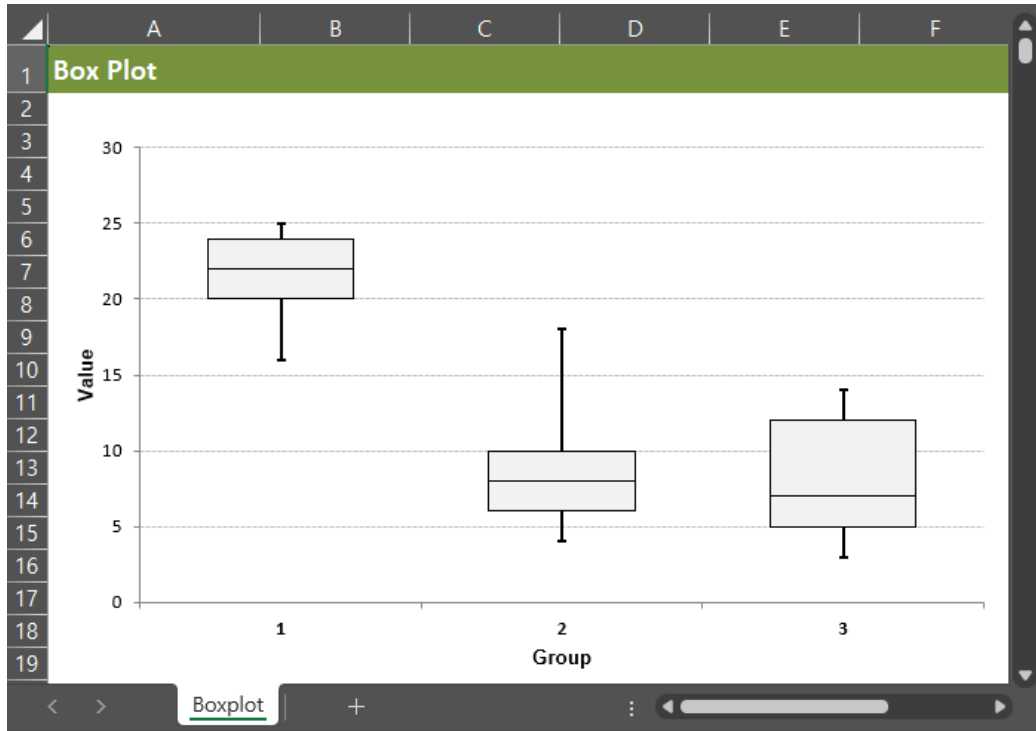
집단명	1	2	3
1			
2		0.0486*	
3		0.0216*	1.0000

\* 표시: 검정확률  $P < 0.05$

Dunn 다중검정에서 Diet 1 vs. 2 ( $p = 0.0486$ ), Diet 1 vs. 3 ( $p = 0.0216$ ) 비교 검정  $p < 0.05$  으로 통계적으로 유의하다.

## 4 Excel Chart

**Box Plot:** Excel 창에서 Box plot 이 나타난다.






## Exercise

### Friedman Test



음악요법이 심리상태에 미치는 영향을 알아보기 위하여 10 명에서 시행 전과 시행 3 개월후, 6 개월 후에 행복지수를 측정하였다. 음악 요법이 효과가 있는지 유의수준 0.05 에서 검정하시오.

	Happiness Score		
	Before	After 3M	After 6 M
1	1	3	3
2	2	4	3
3	3	3	5
4	4	4	8
5	3	2	6
6	2	3	5
7	2	5	6
8	3	7	8
9	3	5	9
10	2	8	10

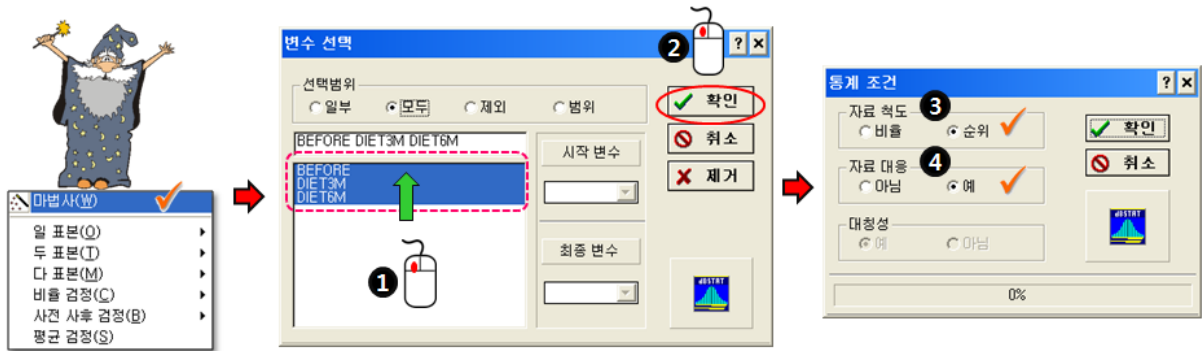


# dBSTAT

Data: Friedman.dbf

1 메뉴 → [통계] ▶ [비교통계] ▶ [마법사] 메뉴를 선택한다.

💡 [통계] ▶ [비교통계] ▶ [다 표본] ▶ [프리드만 검정]을 선택해도 동일 결과가 출력된다.



- ① [변수 선택] 대화상자에서 변수 BEFORE, AFTER3M, AFTER6M 을 선택한다.
- ② [확인] 버튼을 클릭하여 선택 작업을 마치면 [통계 조건] 대화상자가 나타난다.
- ③ [통계 조건] → [자료 척도] ○[순위] ④ [자료 대응] ○[예] 항목을 선택한다. [확인] 버튼을 누른다.

## 2 통계결과

Friedman test

Friedman Two Way ANOVA  
변 수 명: BEFORE 부터 AFTER6M  
표본크기: 10

1 [대표본 근사검정]  
통 계 량:  $\chi^2 = 13.6500$                       동 순위 교 정 후:  $\chi^2 = 14.7568$   
유의수준 ( $\alpha = 0.05$ ):  $\chi^2 = 5.99$                       확 률:  $P = 0.0006$

결 론: 비교 집단 분포(중앙값) 간에 통계적으로 유의한 차이가 있습니다.

2 Friedman 다중비교표

집 단 명	BEFORE	AFTER3M	AFTER6M
BEFORE			
AFTER3M	0.2805		
AFTER6M	0.0006+	0.1326	

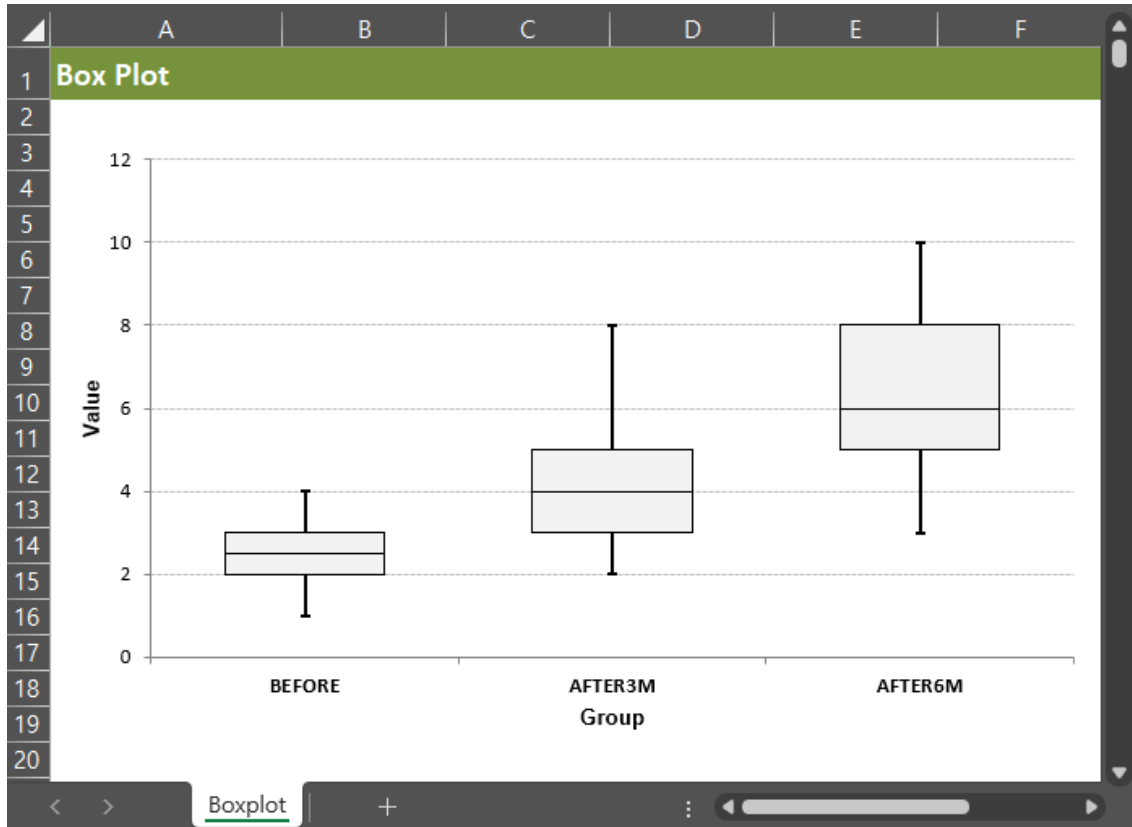
\* 표시: 검정확률  $P < 0.05$

① 대표본 근사검정  $\chi^2 = 14.7568 \Rightarrow p = 0.001$  (0.0006)

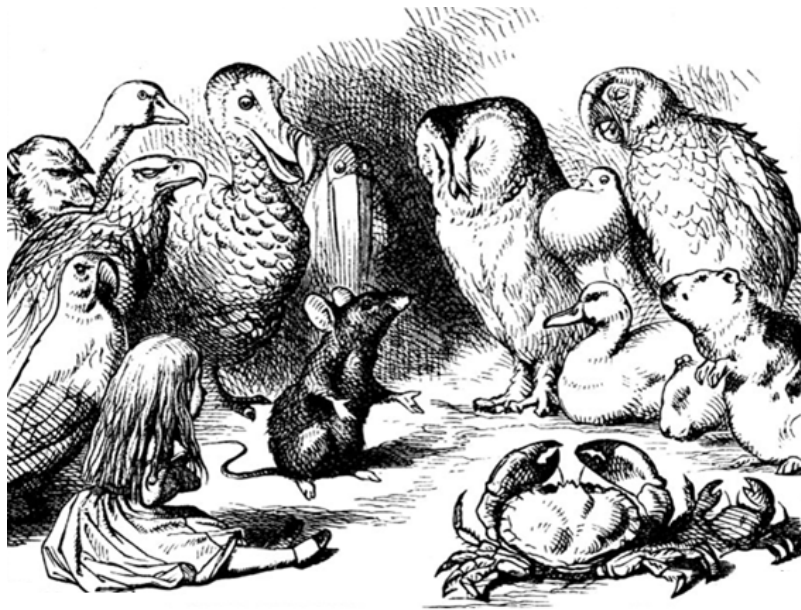
② 사후검정 결과 Before – After 6M 집단 간에 통계적으로 유의한 차이가 있다.

### 3 Excel Chart

**Box Plot:** Excel 창에 Box plot 이 나타난다



# Relationship



*“Found what?” said Duck.*

*“Found It,” the Mouse replied rather crossly:*

*“of course you know what ‘it’ means.”*



## Exercise

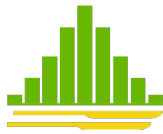
### Pearson Correlation



성인에서 체중(kg)과 키(cm)를 측정하여 다음과 같은 자료를 얻었다.

체중과 키 간에 상관관계가 있는지 유의수준 5%로 검정하시오.

Subject	Weight (kg)	Height (cm)
1	65	162
2	78	165
3	62	168
4	67	165
5	84	165
6	60	167
7	75	155
8	58	170
9	72	152
10	64	175
11	56	152
12	46	150
13	58	158
14	72	148
15	63	161
16	60	165
17	48	155
18	55	158
19	61	160
20	62	162

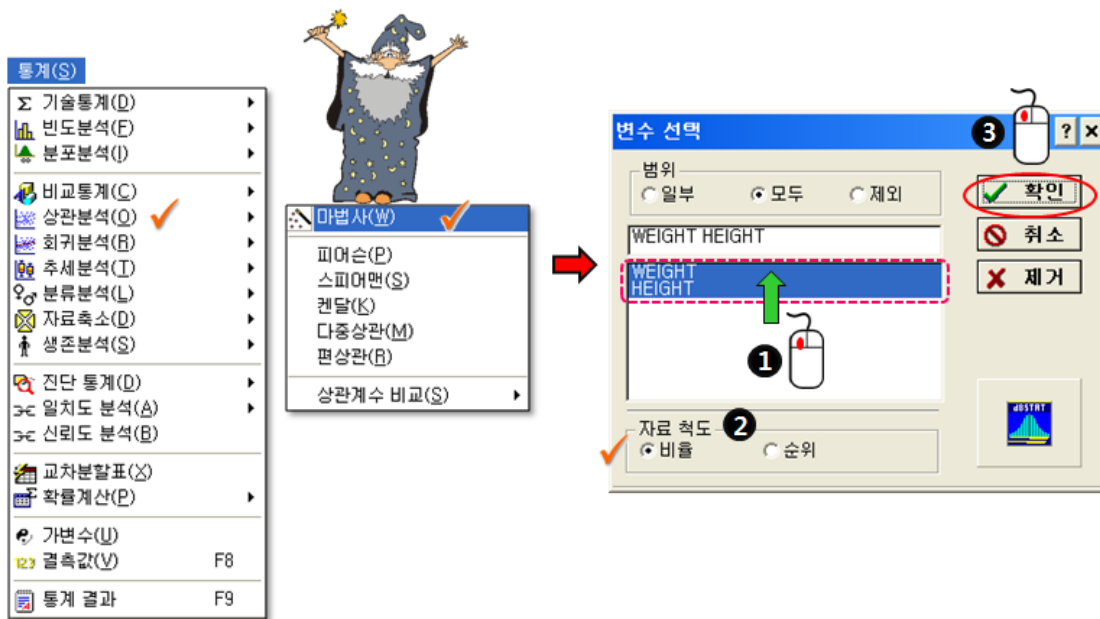


# dBSTAT

Data: corr.dbf

1 메뉴 → [통계] ▶ [상관분석] ▶ [마법사] 메뉴를 선택한다.

💡 [통계] ▶ [상관분석] ▶ [피어슨]을 선택하여도 동일하다.

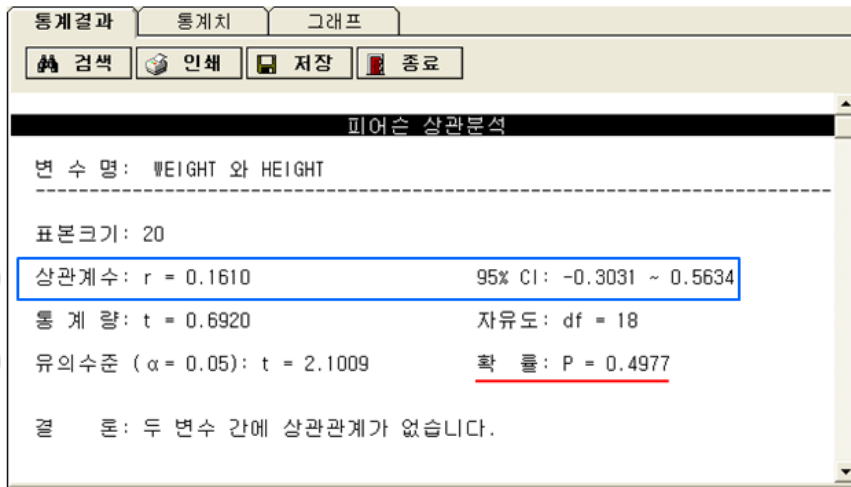


[변수 선택] 대화상자에서 변수 WEIGHT, HEIGHT 를 선택한다.

[자료척도] 선택상자에서 ②[비율] 항목을 선택한다.

[확인] 버튼을 클릭하여 마친다.

## 2 통계결과

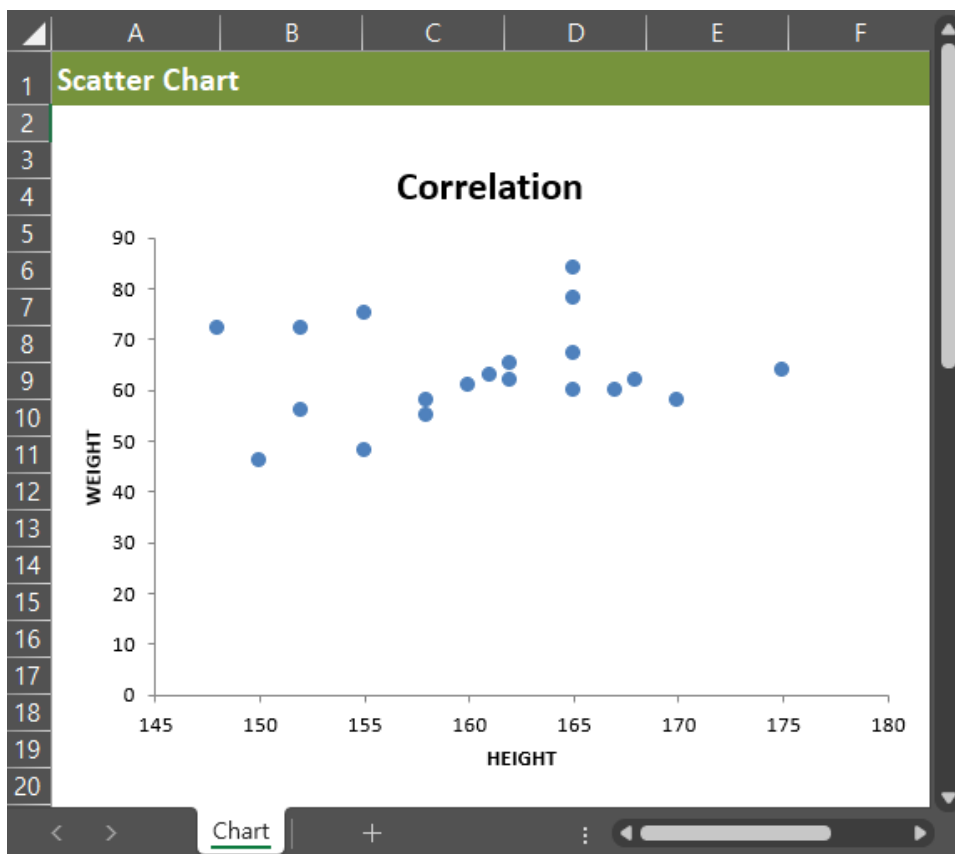


① 상관계수  $r = 0.1610$ , 95% 신뢰구간 =  $(-0.3031, 0.5634)$

② 유의확률  $p = 0.4977 \Rightarrow P > 0.05$  이므로 유의한 상관관계가 없다고 결론을 내린다.

## 3 Excel Chart

**Scatter Plot:** Excel 창에 산점도(Scatter plot)가 나타난다





## Exercise

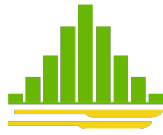
### Spearman Correlation

---



입원 환자 10 명에서 IQ 와 불안지수(anxiety score)가 상관관계가 있는지 유의수준 5%로 검정하시오. (불안지수: 0 ~ 10)

Patient	1	2	3	4	5	6	7	8	9	10
IQ	86	97	99	100	101	103	106	110	112	113
Anxiety Score	1	6	8	7	10	9	3	5	2	4



# dBSTAT

Data: Spearman.dbf

## 상관분석

1 메뉴 → [통계] ▶ [상관분석] ▶ [마법사] 메뉴를 선택한다.



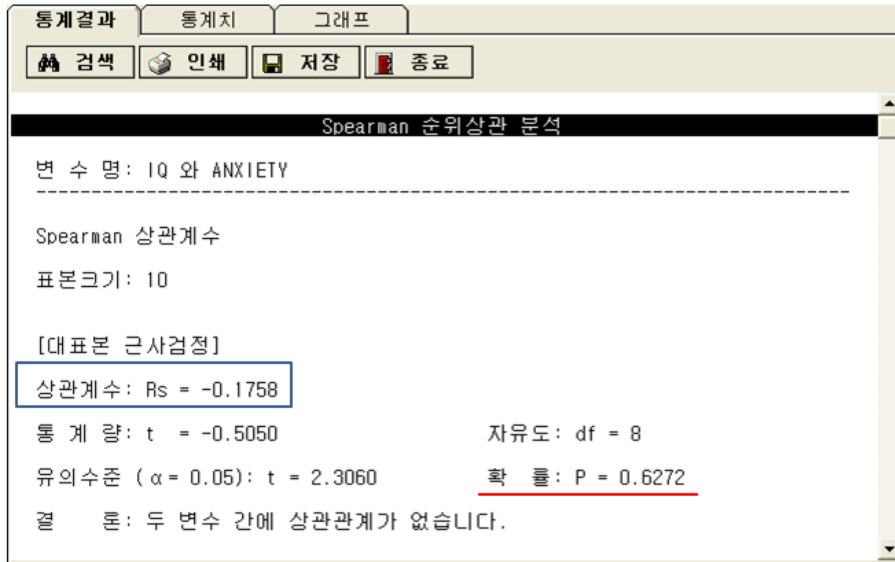
[통계] ▶ [상관분석] ▶ [스피어맨]을 선택하여도 동일하다.

① [변수 선택] 대화상자에서 변수 IQ, ANXIETY 를 선택한다.

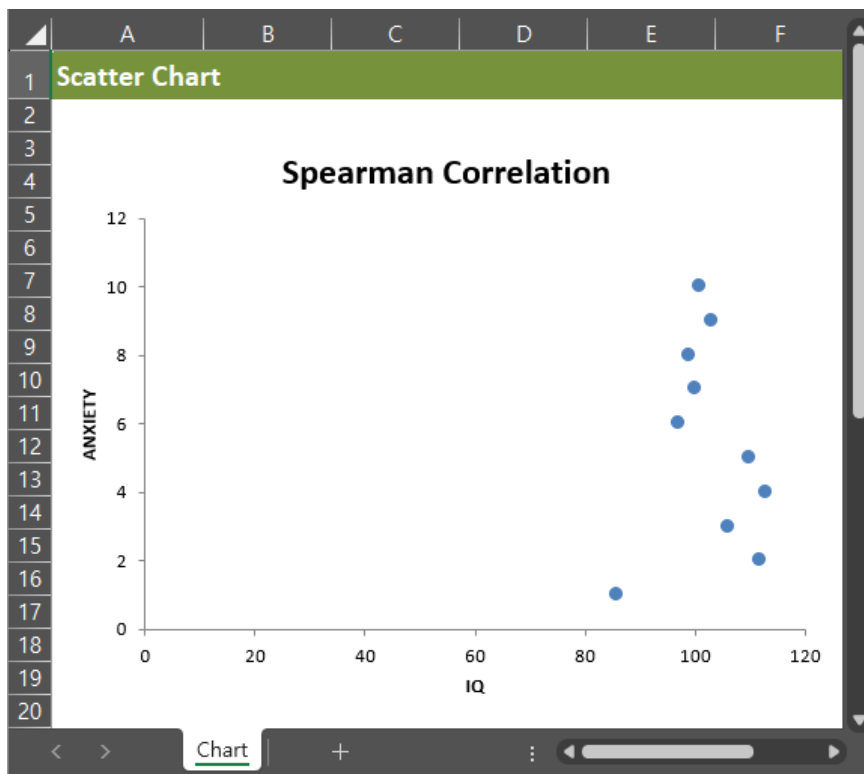
② [자료척도] 선택상자에서 ②[순위] 항목을 선택한다. ③ [확인] 버튼을 클릭하여 마친다.

## 2 통계결과

상관계수  $r_s = -0.1758$ , 유의확률  $p = 0.6272 \Rightarrow P > 0.05$  이므로 유의한 상관관계가 없다.



## 3 Excel Chart





## Exercise

### Partial Correlation

---



심장병 센터를 방문한 성인 환자에서 혈압(mmHg)과 연령, 체중(kg)을 측정하여 다음과 같은 자료를 얻었다. 연령의 영향을 배제하고 체중과 혈압의 상관관계를 구하여 본다.

ID	BP	Age	Weight	ID	BP	Age	Weight
001	158	41	65	011	144	39	56
002	185	60	78	012	138	45	46
003	152	41	62	013	145	47	58
004	159	47	67	014	162	65	72
005	176	66	84	015	142	46	63
006	156	47	60	016	170	67	60
007	184	68	75	017	124	42	48
008	138	43	58	018	158	67	55
009	172	68	72	019	154	56	61
010	168	57	64	020	162	64	62



dBSTAT

Data: mreg.dbf

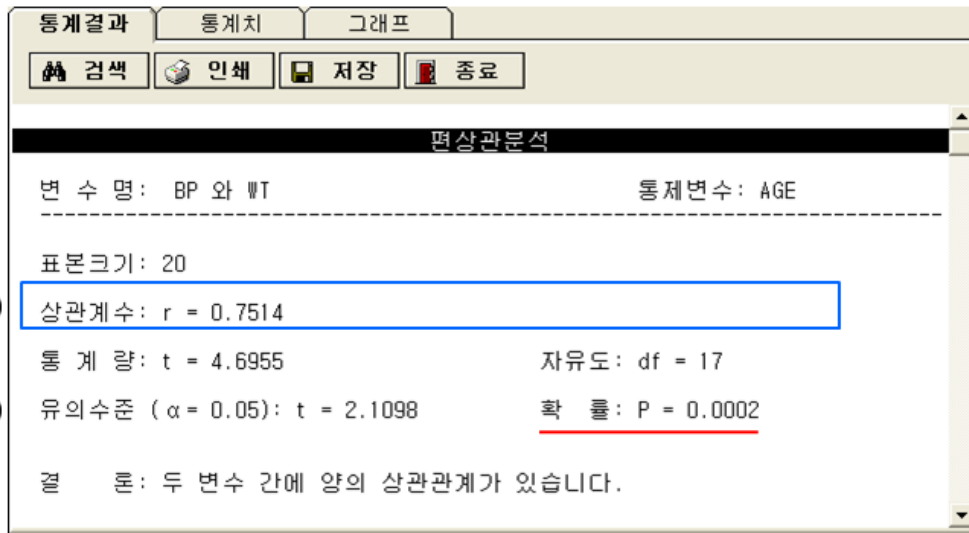
## 편상관분석 (Partial Correlation)

1 메뉴 → [통계] ▶ [상관분석] ▶ [편상관] 메뉴를 선택한다.

The screenshot illustrates the steps to perform a partial correlation analysis in dBSTAT. On the left, the '통계(S)' (Statistics) menu is open, with '상관분석(Q)' (Correlation Analysis) selected. A sub-menu is shown with '편상관(R)' (Partial Correlation) highlighted. A wizard character is positioned between the menu and the dialog box. The '변수 선택' (Variable Selection) dialog box is shown with '자료 변수 1' (Data Variable 1) set to 'BP' and '자료 변수 2' (Data Variable 2) set to 'WT'. The '통제 변수' (Control Variable) is set to 'AGE'. The '확인' (OK) button is circled in red, and a mouse cursor is pointing at it. The steps are numbered 1, 2, and 3.

- ① [변수 선택] 상자에서 [자료 변수 1] ▷ BP [자료 변수 2] ▷ WT 를 입력한다.
- ② [통제 변수] ▷ AGE 를 혼선택변수로 입력한다.
- ③ [확인] 버튼을 클릭하여 선택 작업을 마치면 [통계결과] 창이 나타난다.

## 2 통계결과

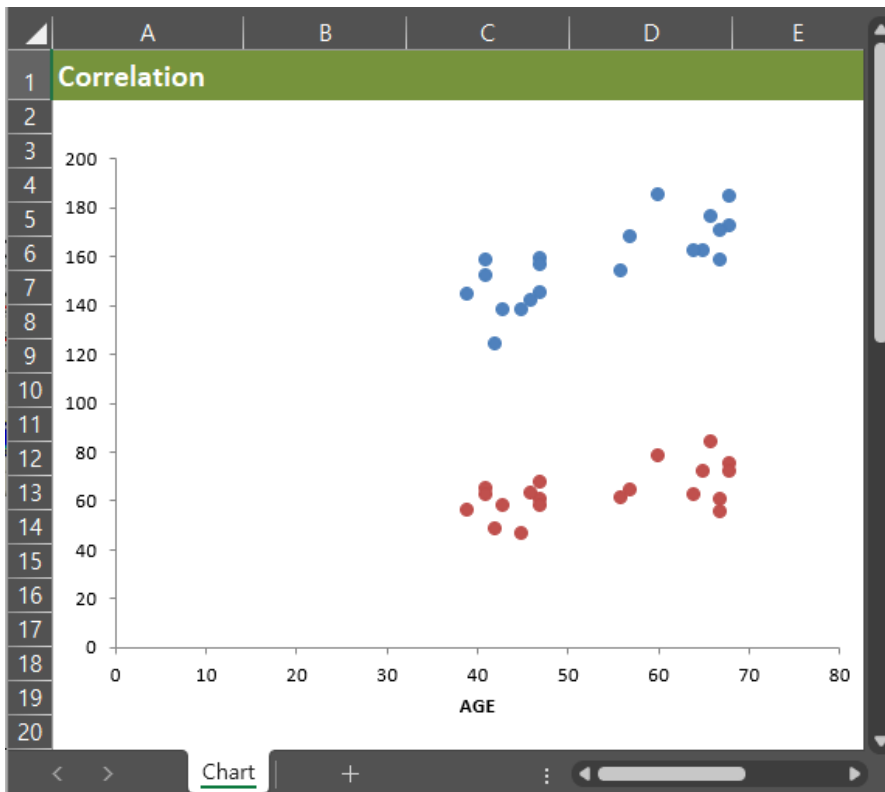


① 편상관계수  $r = 0.7514$

②  $P = 0.0002 < 0.05$  이므로 통계적으로 유의한 상관관계가 있다.

## 3 Excel Chart

**Scatter Plot:** Excel 창에 산점도(Scatter plot)가 나타난다



# Regression





## Exercise

### Regression Analysis



심장병 센터를 방문한 성인 환자에서 혈압(mmHg)과 연령을 측정하여 다음과 같은 자료를 얻었다. 연령으로 혈압을 예측할 수 있는지 분석하여 본다.

Subject	BP (mmHg)	Age (year)
1	158	41
2	185	60
3	152	41
4	159	47
5	176	66
6	156	47
7	184	68
8	138	43
9	172	68
10	168	57
11	144	39
12	138	45
13	145	47
14	162	65
15	142	46
16	170	67
17	124	42
18	158	67
19	154	56
20	162	64



dBSTAT

Data: sreg.dbf

## 회귀분석

1 메뉴 → [통계] ▶ [회귀분석] ▶ [단순회귀] ▶ [회귀분석] 메뉴를 선택한다.

The image shows the dBSTAT software interface. On the left is the '통계(S)' (Statistics) menu, which is expanded to show '회귀분석(B)' (Regression Analysis) with a checkmark. Below it, a sub-menu is shown with '단순회귀(S)' (Simple Regression) selected. A red arrow points from this menu to the '변수 선택' (Variable Selection) dialog box on the right. The dialog box has two dropdown menus: '독립 변수' (Independent Variable) set to 'AGE' and '종속 변수' (Dependent Variable) set to 'BP'. A mouse cursor is shown clicking on the 'AGE' dropdown, labeled with a circled '1'. The '확인' (OK) button is circled in red and labeled with a circled '2'. The '취소' (Cancel) button is also visible. A progress bar at the bottom shows 0%.

- ① [변수 선택] 대화상자에서 [독립변수] > AGE, [종속변수] > BP 를 입력한다.
- ② [확인] 버튼을 클릭하여 선택 작업을 마치면 [통계결과] 창이 나타난다

## 2 통계결과

통계결과    통계치    그래프

🔍 검색    📄 인쇄    💾 저장    🛑 종료

**단순회귀분석**

① 종속변수: BP                      독립변수: AGE

---

변인	자유도	제곱합	분산	분산비
회귀	1	2804.60	2804.60	24.967
오차	18	2021.95	112.33	

유의수준 ( $\alpha = 0.05$ ):  $F = 4.41$                       확률:  $P = 0.0001$

결론: 두 변수 간에 유의한 선형회귀 관계가 있습니다.

② [회귀방정식]

종속변수(Y): BP

독립변수(X)	회귀계수(B)	표준화 회귀계수	표준오차	t 값	유의성
AGE	1.113	0.762	0.223	4.997	<u>0.0001*</u>

Y 절편 (C) = 97.460

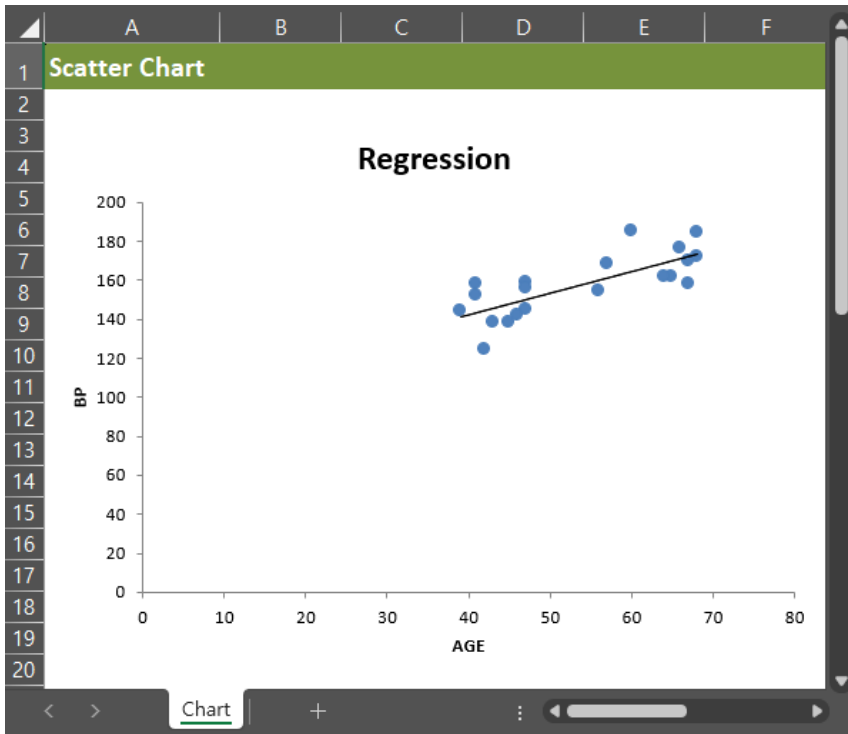
회귀방정식:  $BP(Y) = 1.113 \times AGE(X) + 97.460$

③ 결정계수:  $R^2 = 0.581$

해석: 회귀방정식으로 종속변수의 변동을 58.1% 설명할 수 있습니다.

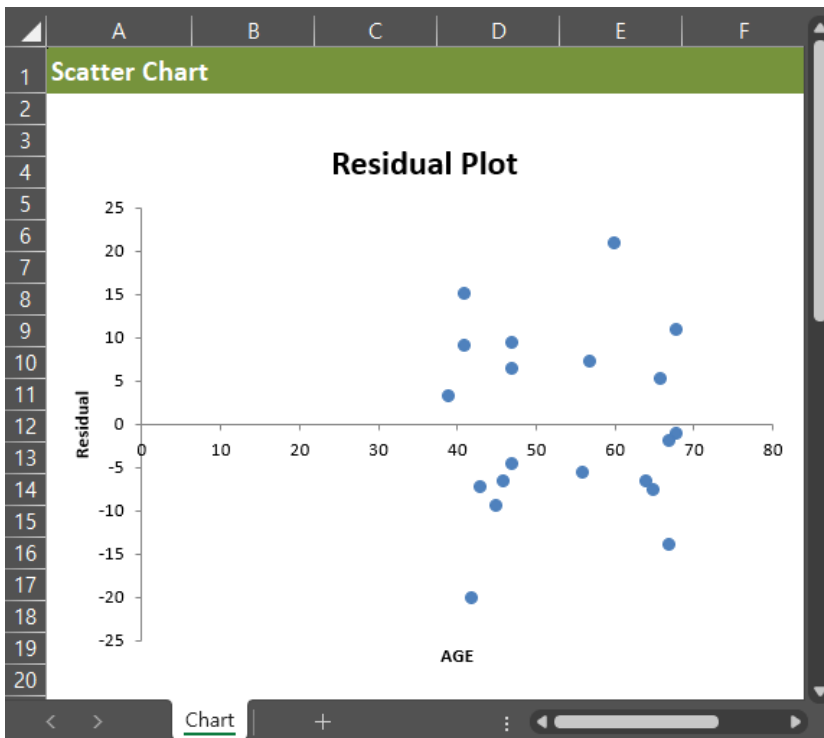
- ① 회귀식에 대한  $p = 0.0001$  이므로 Age 와 BP 는 유의한 선형 회귀관계가 있다.
- ② 개별 독립변수 AGE 에 대한 유의성 검정에서  $p = 0.0001$  이므로 통계적으로 유의하다.
- ③ 결정계수  $R^2 = 0.581$  로 회귀식에 의하여 BP 변동의 약 58%를 설명할 수 있다.

### 3 Excel Chart



### 4 회귀진단

메뉴 → [통계] ▶ [회귀분석] ▶ [단순회귀] ▶ [회귀진단] 메뉴를 선택한다.  
회귀진단 결과가 해석과 함께 출력되며 아래와 같은 잔차그림을 보여준다.





## Exercise

### Multiple Regression Analysis



심장병 센터를 방문한 성인 환자에서 혈압(mmHg), 연령, 체중, 콜레스테롤을 측정하였다. 혈압을 예측하기 위한 가장 적합한 회귀식을 만들어본다.

Subject	BP (mmHg)	Age (year)	Weight (kg)	Cholesterol (mg/dL)
1	158	41	65	244
2	185	60	78	280
3	152	41	62	220
4	159	47	67	261
5	176	66	84	310
6	156	47	60	267
7	184	68	75	290
8	138	43	58	172
9	172	68	72	198
10	168	57	64	181
11	144	39	56	215
12	138	45	46	166
13	145	47	58	221
14	162	65	72	285
15	142	46	63	218
16	170	67	60	260
17	124	42	48	184
18	158	67	55	214
19	154	56	61	221
20	162	64	62	267



dBSTAT

Data: mreg.dbf

## 다중선형회귀분석 (Multiple Linear Regression)

1 메뉴 → [통계] ▶ [회귀분석] ▶ [다중회귀] ▶ [회귀분석] 메뉴를 선택한다.



- ① [통계 조건] 상자에서 [종속변수] > BP [독립변수] > AGE, WT, CHOL 를 입력한다.
- ② [확인] 버튼을 클릭하여 선택 작업을 마치면 [통계결과] 창이 나타난다.

## 2 통계결과

통계결과    통계치    그래프

🔍 검색    🌐 인쇄    💾 저장    🛑 종료

**단순회귀분석**

① 종속변수: BP                      독립변수: AGE

---

변인	자유도	제곱합	분산	분산비
회귀	1	2804.60	2804.60	24.967
오차	18	2021.95	112.33	

유의수준 ( $\alpha = 0.05$ ):  $F = 4.41$                       확률:  $P = 0.0001$

결론: 두 변수 간에 유의한 선형회귀 관계가 있습니다.

② [회귀방정식]

종속변수(Y): BP

독립변수(X)	회귀계수(B)	표준화 회귀계수	표준오차	t 값	유의성
AGE	1.113	0.762	0.223	4.997	<u>0.0001*</u>

Y 절편 (C) = 97.460

회귀방정식:  $BP(Y) = 1.113 \times AGE(X) + 97.460$

③ 결정계수:  $R^2 = 0.581$

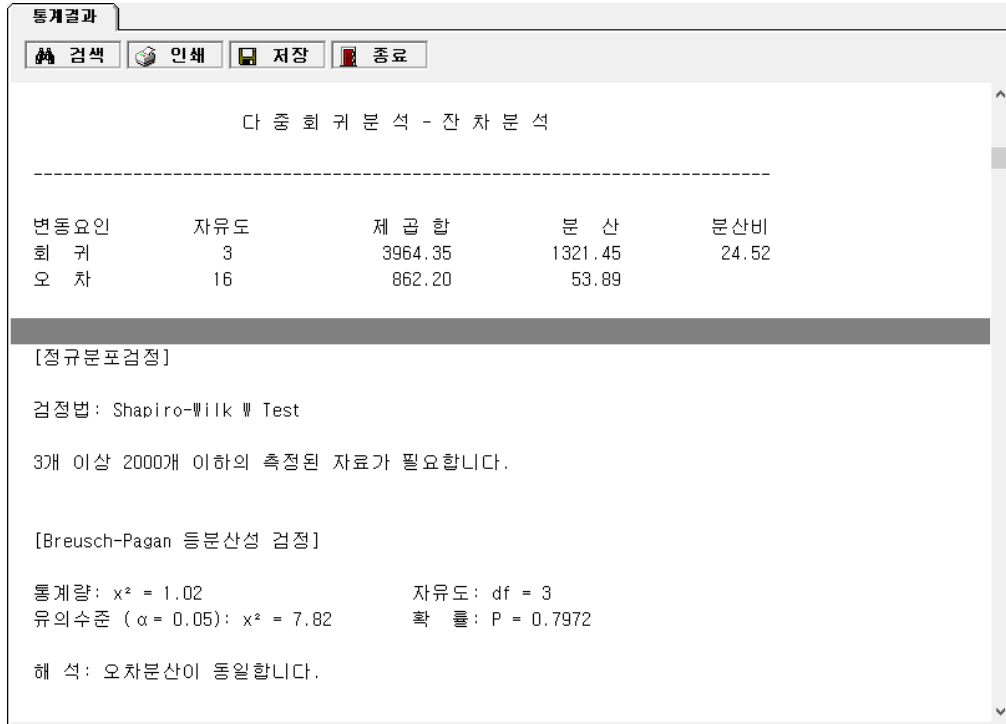
해석: 회귀방정식으로 종속변수의 변동을 58.1% 설명할 수 있습니다.

- ① 회귀식에 대한  $p = 0.0000$  이므로 독립변수들과 BP 는 유의한 선형 회귀관계가 있다.
- ② 개별 독립변수에 대한 유의성 검정에서 AGE, WT ( $p < 0.05$ )가 통계적으로 유의하다.
- ③ 결정계수  $R^2 = 0.821$  로 회귀식에 의하여 혈압(BP) 변동의 약 82%를 설명할 수 있다.

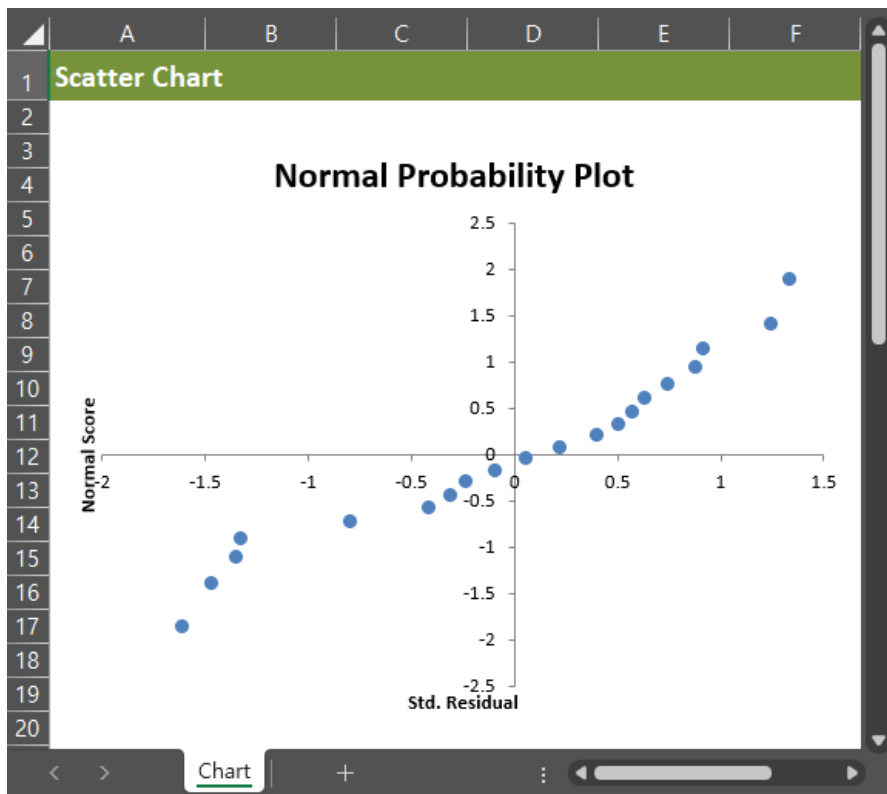


## 4 회귀진단

[통계] ▶ [회귀분석] ▶ [다중회귀] ▶ [회귀진단] 메뉴를 선택한다.



회귀진단 결과가 해석과 함께 출력되며 아래와 같은 정규확률 그림을 보여준다.





## Exercise

### Logistic Regression



심장병 센터를 방문한 환자들을 대상으로 관상동맥심질환(CHD)이 유무에 따라 CHD 위험인자로서 연령, 성별, 체중, 흡연 과거력을 조사하였다. 위험인자들이 CHD 발병과 연관이 있는지 알아보자.

No.	CHD	Age	No.	CHD	Age	No.	CHD	Age
1	0	22	12	0	40	23	0	54
2	0	23	13	1	41	24	1	55
3	0	24	14	0	46	25	1	58
4	0	27	15	0	47	26	1	60
5	0	28	16	0	48	27	0	60
6	0	30	17	1	49	28	1	62
7	0	30	18	0	49	29	1	65
8	0	32	19	1	50	30	1	67
9	0	33	20	0	51	31	1	71
10	1	35	21	1	51	32	1	77
11	0	38	22	0	52	33	1	81

위의 자료의 변수명에 대한 설명과 코드는 다음 Table 에 정리하였다.

No.	변수명	설명	Code/Value
1	CHD	관상동맥심질환 유무	0=없음; 1=있음
2	Age	연령 (세)	
3	Sex	성별	1=여자; 2=남자
4	Weight	체중 (kg)	
5	Smoking	흡연 과거력	1=비흡연; 2=흡연

로지스틱 모형을 개발하고 통계적 의의가 있는지 판단하시오.



로지스틱 회귀분석의 전제조건에 대한 검정은 회귀분석과 유사하다.



**dBSTAT**

Data: Logistic.dbf

## 로지스틱 회귀분석

1 메뉴 → [통계] ▶ [회귀분석] ▶ [로지스틱회귀] ▶ [회귀분석] 메뉴를 선택한다.

The image shows the dBSTAT software interface. On the left, the '통계(S)' menu is open, and the path '통계(S) > 회귀분석(B) > 로지스틱회귀(L) > 회귀분석(B)' is highlighted. A wizard character is shown next to the menu items. In the center, a red arrow points to the '통계 조건' dialog box. The dialog box has three sections: '종속 변수' (Dependent Variable) with 'CHD' selected, '독립 변수' (Independent Variables) with 'AGE WT' selected, and '집단 변수' (Covariate) with 'SEX SMOKING' selected. The '확인' (OK) button is circled in red, and a mouse cursor is shown clicking it. The '확인' button is also circled in red. The '확인' button is circled in red. The '확인' button is circled in red.

- ① [통계 조건] 상자에서 [종속변수] ▷ CHD [독립변수] ▷ AGE, WT 를 입력한다.
- ② [통계 조건] 상자에서 [집단변수] ▷ SEX, SMOKING 을 입력한다.
- ③ [확인] 버튼을 누른다.

## 2 통계결과

통계결과    통계치    그래프

🔍 검색    🌐 인쇄    💾 저장    🛑 종료

로지스틱 회귀분석

① [ 로지스틱 회귀모형의 적합성 ]

-2 Log Likelihood : 19.26

Chi-square      자유도      확률  
25.732            4            0.0000

결론: 로지스틱 회귀분석 모형이 적합합니다.

② [ 로지스틱 회귀방정식 ]

LOGIT CHD =  $B_0 + B_1 \times X_1 + B_2 \times X_2 + \dots + B_n \times X_n$   
( $B_0$ : 상수,  $B_n$ : 회귀계수,  $X_n$ : 독립변수)

종속변수 : CHD = 1      (Reference category = 0)

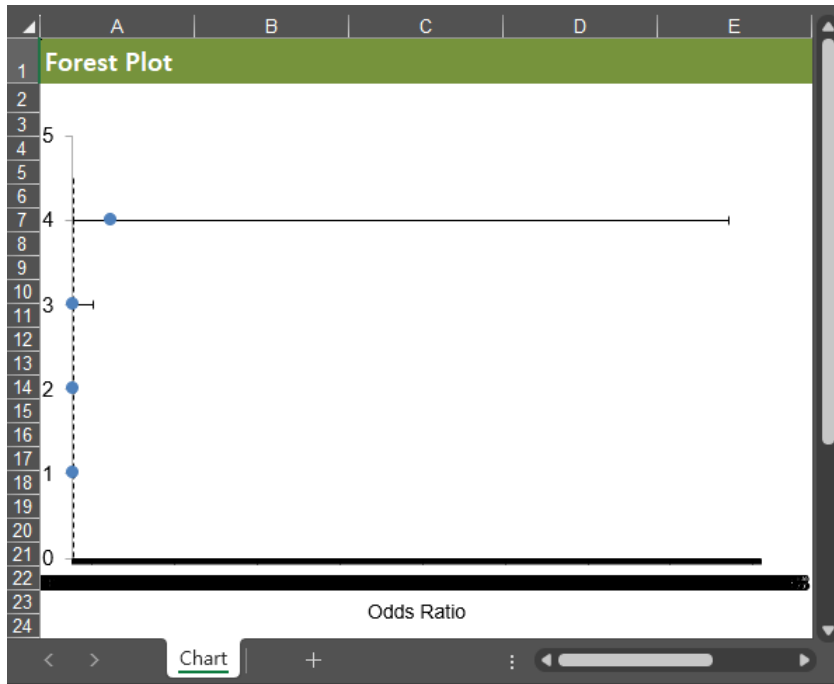
독립변수(X)	회귀계수(B)	표준오차(SE)	Wald	유의성	편상관	표준화 회귀계수	승산비(OR)
Constant	-12.467						
AGE	0.138	0.069	3.999	0.046*	0.211	0.135	1.148
WT	0.061	0.077	0.624	0.430	0.000	0.000	1.063
SEX(1)	0.387	1.259	0.095	0.758	0.000	0.000	1.473
SMOKING(1)	3.485	1.433	5.917	0.015*	0.295	0.144	32.625

결론: \* 표시(P < 0.05) 된 공변수가 통계적으로 의미가 있습니다.

- ① 회귀식에 대한  $p = 0.0000$  이므로 로지스틱 회귀모형이 적합하다.
- ② 개별 독립변수의 회귀계수, 유의성, 승산비를 보여준다.
- ③ AGE, SMOKING 변수가 통계적으로 유의하다 ( $p < 0.05$ ).

### 3 Excel Chart

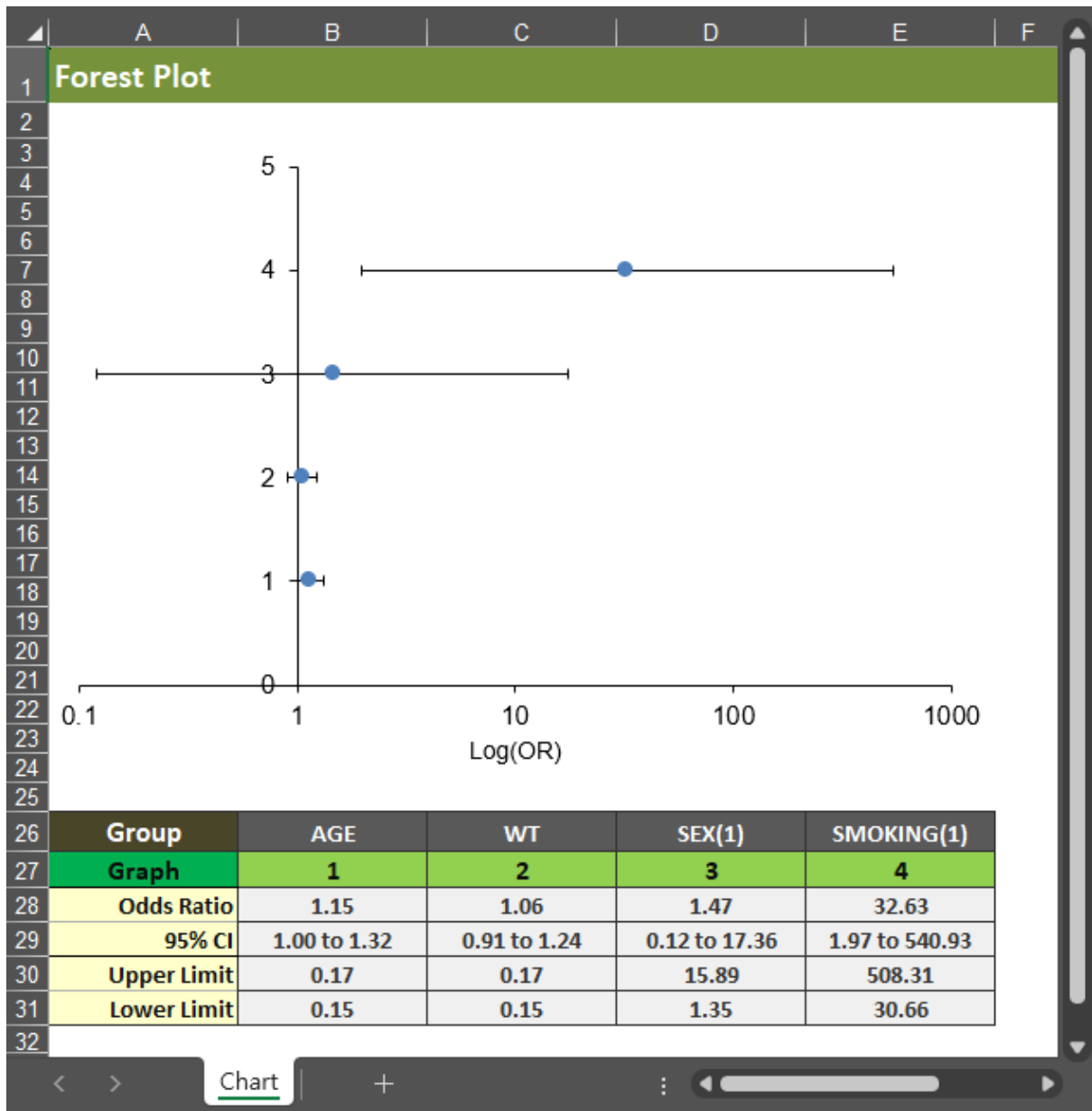
**Forest Plot:** 통계결과 창에서 다음과 같은 숲 그림이 나타난다.



X 축에서 마우스 오른쪽 버튼 클릭 ▶ 선택메뉴 옵션에서 [축 서식] ▶ [로그 눈금 간격]

The image shows the process of applying a logarithmic scale to the x-axis of the Forest Plot chart. It starts with a right-click on the x-axis, followed by selecting '축 서식(F)...' from the context menu. This opens the '축 서식' (Axis Format) task pane. In the '축 옵션' (Axis Options) section, the '로그 눈금 간격' (Log Scale Interval) checkbox is checked, and the '기준(B)' (Base) is set to 10.

X 축 눈금 간격 ▶ 로그 scale 로 바꾸고 X 축 label ▶ "Log(OR)"로 바꾼다.



## 승산비

Table 에서 개별 독립변수의 승산비와 95% 신뢰구간을 보여준다.

## 4 회귀진단

[통계] ▶ [회귀분석] ▶ [로지스틱회귀] ▶ [회귀진단] 메뉴를 선택한다.

회귀진단 결과가 해석과 함께 출력되며 아래와 같은 정규확률 그림을 보여준다.

통계결과

🔍 검색 🔄 인쇄 💾 저장 🛑 종료

[이상값]

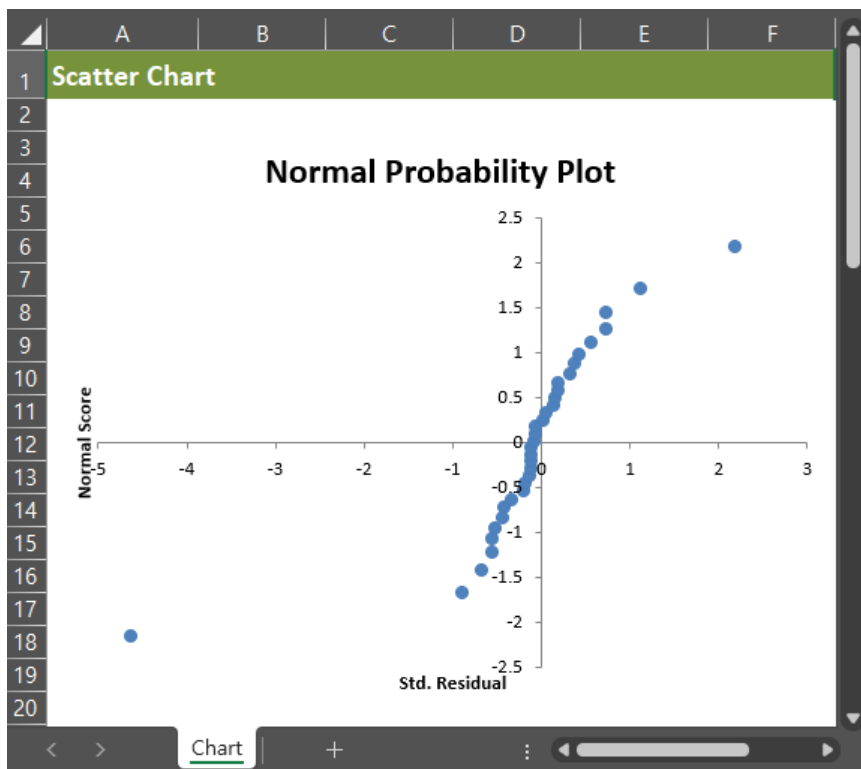
검정법: Standardized Residual

해 석: 기각치 2 보다 절대값이 큰 관찰점(\*)이 이상값입니다.

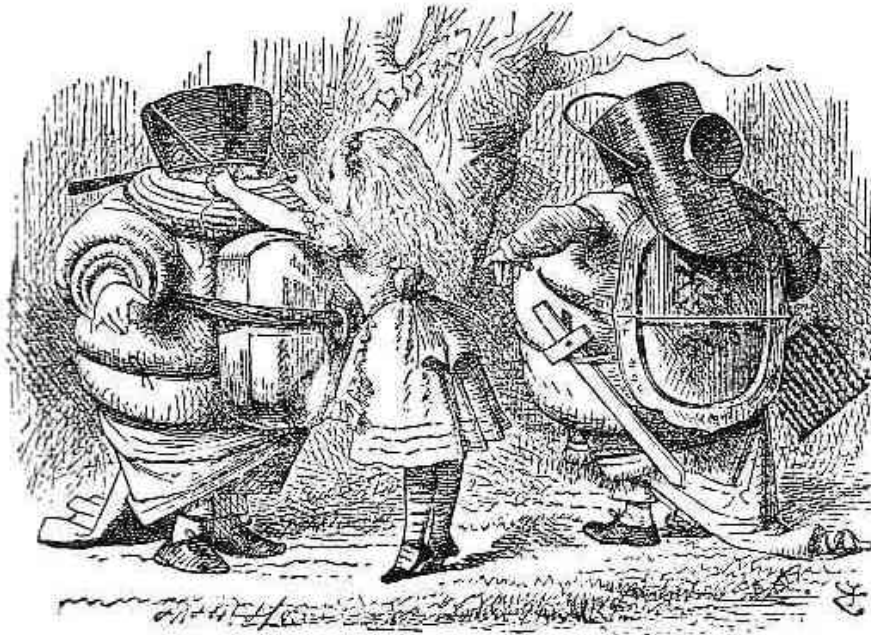
[다중공선성]

변수명	분산팽창인자
AGE	1.3155
WT	1.3155

해 석: 분산팽창인자가 10 보다 큰 변수(\*)에서 다중공선성이 있습니다.



# Association





## Exercise

### Chi-square Test



성인에서 흡연(smoking)과 천식(asthma)의 연관성을 알아보는 후향적 연구를 하였다. 천식 환자에서 흡연 과거력을 조사하였다. 흡연과 천식이 연관성이 있는지 유의수준 0.05 에서 통계적 검정을 하시오.

No.	Smoking	Asthma	No.	Smoking	Asthma	No.	Smoking	Asthma
1	1	1	21	1	2	41	2	2
2	1	1	22	1	2	42	2	2
3	1	1	23	1	2	43	2	2
4	1	1	24	1	2	44	2	2
5	1	1	25	1	2	45	2	2
6	1	1	26	1	2	46	2	2
7	1	1	27	1	2	47	2	2
8	1	1	28	1	2	48	2	2
9	1	1	29	1	2	49	2	2
10	1	1	30	1	2	50	2	2
11	1	1	31	2	1	51	2	2
12	1	1	32	2	1	52	2	2
13	1	1	33	2	1	53	2	2
14	1	1	34	2	1	54	2	2
15	1	1	35	2	1	55	2	2
16	1	1	36	2	1	56	2	2
17	1	1	37	2	1	57	2	2
18	1	1	38	2	1	58	2	2
19	1	1	39	2	1	59	2	2
20	1	1	40	2	1	60	2	2

smoking: 1=yes , 2=no    asthma : 1= yes , 2=no



dBSTAT

Data: chi-square.dbf

## Chi-square 검정

1 메뉴 → [통계] ▶ [비교통계] ▶ [마법사]를 선택하면 [변수 선택] 대화상자가 나타난다.

💡 [통계] ▶ [비교통계] ▶ [두 표본] ▶ [카이제곱 검정]을 선택해도 동일한 결과가 출력된다.

The screenshot shows the '통계(S)' menu with '비교통계(C)' selected. A sub-menu '마법사(W)' is open, showing options like '일 표본(O)', '두 표본(I)', and '비율 검정(C)'. A red arrow points to the '변수 선택' dialog box. In this dialog, 'SMOKING ASTHMA' is selected in the variable list, and the '확인' button is circled in red. A mouse cursor icon with a '1' is shown clicking on the variable, and another with a '2' is shown clicking on the '확인' button.

- ① [변수 선택] 대화상자에서 변수 SMOKING, ASTHMA 를 선택한다.
- ② [확인] 버튼을 클릭하여 선택 작업을 마치면 [통계 조건] 대화상자가 나타난다.

The '통계 조건' dialog box is shown. Under '자료 대응', the '아님' radio button is selected. The '확인' button is circled in red. A mouse cursor icon with a '1' is shown clicking on the '아님' option, and another with a '2' is shown clicking on the '확인' button.

- ① [통계 조건] 대화상자에서 [자료 대응] Ⓞ[아님] 항목을 선택한다.
- ② [확인] 버튼을 클릭하여 선택 작업을 마치면 통계결과가 출력된다.

## 2 통계결과

dBSTAT에서는 [통계 조건]에 따라서 [마법사]가 적절한 통계검정을 시행한다.

통계결과 | 통계치 | 그래프

🔍 검색 | 🌐 인쇄 | 💾 저장 | 🛑 종료

**1** 교차분할표

관찰값: 수/퍼센트(총)/퍼센트(행)/퍼센트(열)

행   열	ASTHMA		합 계
	1	2	
SMOKING 1	20	10	30
	33.33	16.67	
	66.67	33.33	
	66.67	33.33	
2	10	20	30
	16.67	33.33	
	33.33	66.67	
	33.33	66.67	
합 계	30	30	60

**2** [ Chi-Square Test ]

기대빈도가 5미만인 칸(cell) 수: 0 ( 0% )  
 기대빈도가 1미만인 칸(cell) 수: 0

**3** 통계량:  $\chi^2 = 5.40$       자유도:  $df = 1$   
 유의수준 ( $\alpha = 0.05$ ):  $\chi^2 = 3.84$       확률:  $P = 0.0201$

결론: 두 변수 (열과 행) 간에 유의한 연관성이 있습니다.

- ① SMOKING을 행, ASTHMA를 열로 한 교차분할표를 보여준다.
- ② 기대빈도 < 5 인 칸(cell)이 0으로 없으므로 카이제곱 전제조건을 만족한다.
- ③ 유의성 검정

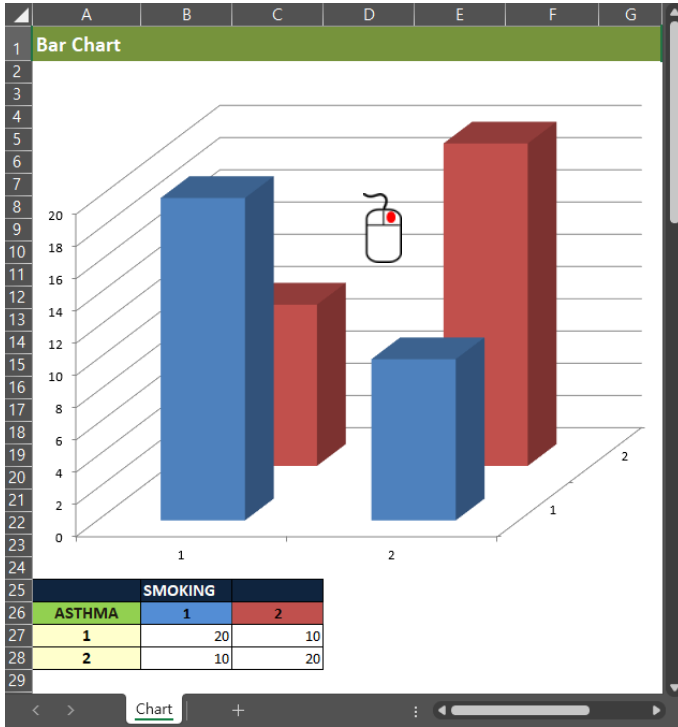
$\chi^2 = 5.40$ , 유의수준  $\alpha = 0.05$ ,  $df = 1 \Rightarrow$  양측검정 유의확률  $p = 0.02$

### [해석]

smoking과 asthma 두변수는 통계적으로 유의한 연관성이 있다.

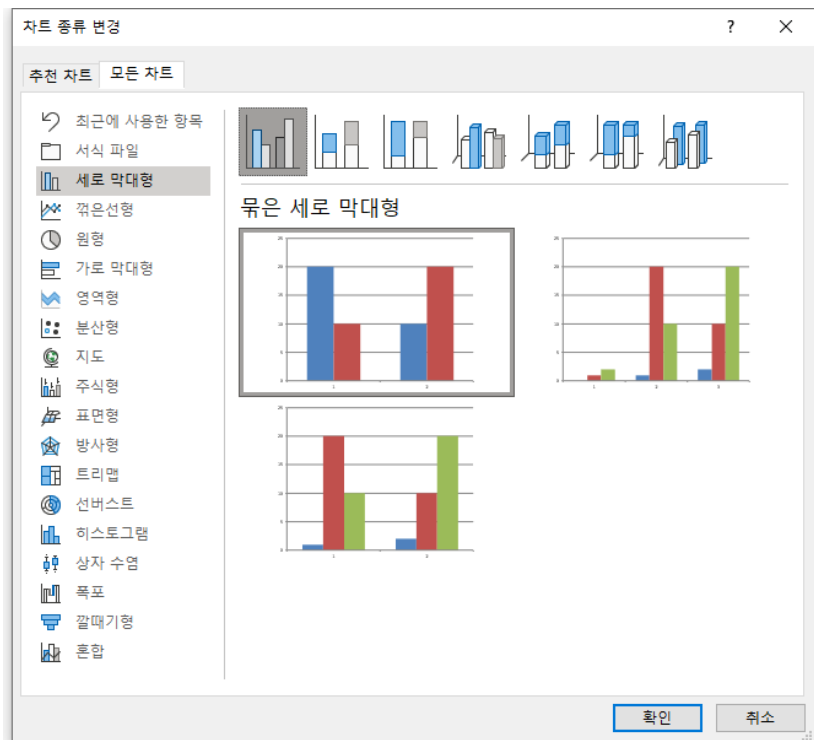
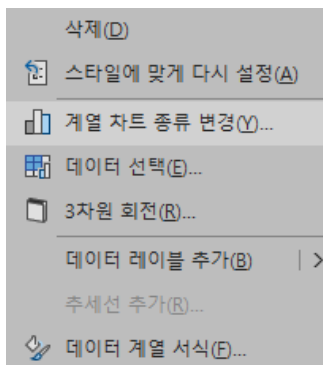
### 3 Excel Chart

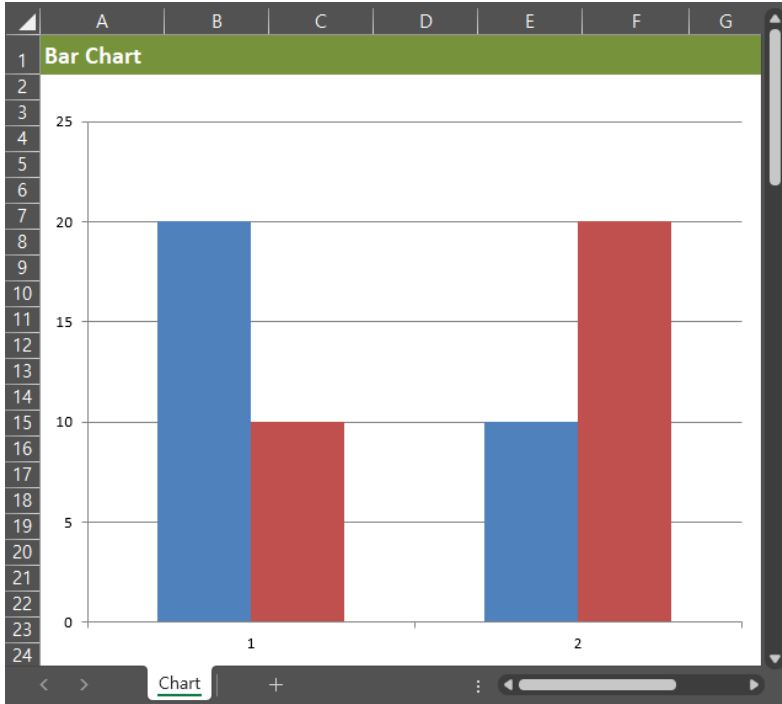
**Bar Chart:** 다음과 같이 3D 막대그래프가 나타난다.



그래프 창에서 마우스 오른쪽 버튼 클릭 ▶ 선택메뉴 옵션에서 **[계열 차트 종류 변경]**

▶ **[세로 막대형]** 선택하면 2D 그래프가 나타난다.





## Graphic Presentation



3D 그래프는 슬라이드 발표할 때에 사용하면 보기 좋으나 통계값 표현이 정확하지 못하므로 학술 논문을 출판할 때에는 2D 그래프가 올바른 방법이다.



## Exercise

### Fisher's Exact Test

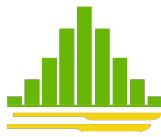
---



성인에서 흡연(smoking)과 천식(asthma)의 연관성을 알아보는 후향적 연구를 하였다. 천식 환자에서 흡연 과거력을 조사하였다. 흡연과 천식이 연관성이 있는지 유의수준 0.05 에서 통계적 검정을 하시오.

No.	Smoking	Asthma	No.	Smoking	Asthma	No.	Smoking	Asthma
1	1	1	7	1	1	13	2	1
2	1	1	8	1	1	14	2	2
3	1	1	9	1	2	15	2	2
4	1	1	10	2	1	16	2	2
5	1	1	11	2	1	17	2	2
6	1	1	12	2	1	18	2	2

smoking: 1=yes , 2=no    asthma : 1= yes , 2=no



# dBSTAT

Data: Fisher.dbf

## 피셔 정확검정 (Fisher's Exact test)

1 메뉴 → [통계] ▶ [비교통계] ▶ [마법사] 메뉴를 선택한다.

dBSTAT에서는 [통계 조건]에 따라서 [마법사]가 적절한 통계검정을 시행한다.



[통계] ▶ [비교통계] ▶ [두 표본] ▶ [피셔 정확확률검정]도 동일한 결과가 출력된다.

## 2 통계결과

The screenshot shows the '통계 결과' (Statistical Results) window. It contains a contingency table for '교차분할표' (Contingency Table) and the results of a 'Fisher Exact Test'.

**교차분할표**

관찰값: 수/퍼센트(총)/퍼센트(행)/퍼센트(열)

행 \ 열	ASTHMA		합계
	1	2	
SMOKING	1	1	9
	81	5.56	
	44.44	11.11	
	88.89	16.67	
2	5	9	
4	27.78		
22.22	55.56		
33.33	83.33		
합계	12	6	18

**Fisher Exact Test**

기대빈도가 5미만인 칸(cell) 수: 2  
기대빈도가 1미만인 칸(cell) 수: 0

FISHER 정확 확률: 하단 확률 = 0.0656 상단 확률 = 0.0656  
단측 검정: P = 0.0656 양측 검정: P = 0.1312

결론: 두 변수 (열과 행) 간에 유의한 연관성이 없습니다.

- ① SMOKING 을 행, ASTHMA 를 열로 한 교차분할표를 보여준다.
- ② 기대빈도 < 5 인 칸(cell)의 수가 2 로 카이제곱 전제조건을 만족하지 않는다.
- ③ 유의성 검정: 양측검정 유의확률  $p = 0.1312$

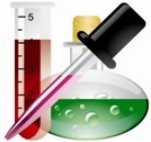
### [해석]

smoking 과 asthma 두변수는 통계적으로 유의한 연관성이 있다.



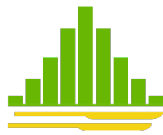
## Exercise

### McNemar's Test



질병의 진단에 있어서 진단법이 차이가 있는지 30 명을 대상으로 두 가지 방법으로 검사하여 양성과 음성으로 나누어 다음 표를 작성하였다. 두 가지 진단법이 일치하는지 유의수준 0.05 로 검정하시오.

		진단법 2		
		양성	음성	합계
진단법 1	양성	4	0	4
	음성	12	14	26
	합계	16	14	30



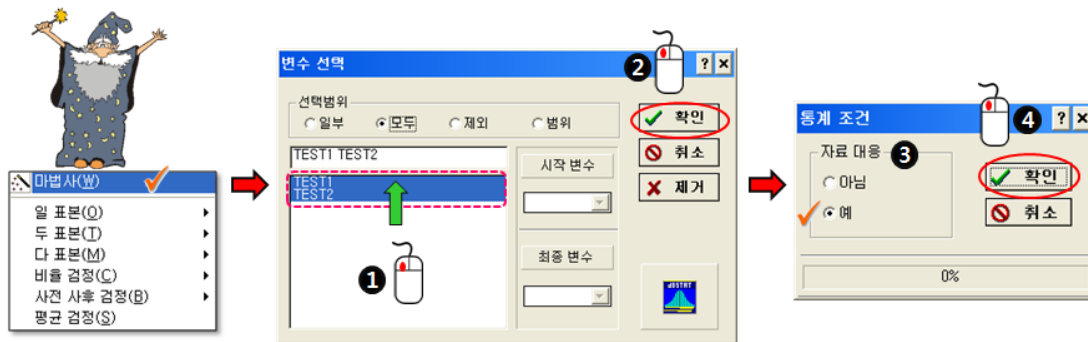
# dBSTAT

Data: McNemar.dbf

## 맥네마 검정 (McNemar's test)

1 메뉴 → [통계] ▶ [비교통계] ▶ [마법사] 메뉴를 선택한다.

💡 [통계] ▶ [비교통계] ▶ [두 표본] ▶ [맥네마 검정]도 동일한 결과가 출력된다.



- ① [변수 선택] 대화상자에서 변수 TEST1, TEST2 를 선택한다.
- ② [확인] 버튼을 클릭하여 선택 작업을 마치면 [통계 조건] 대화상자가 나타난다.
- ③ [통계 조건] 대화상자에서 [자료 대응]  [예] 항목을 선택한다.
- ④ [확인] 버튼을 클릭하여 선택 작업을 마치면 통계결과가 출력된다.

McNemar 검정법

관찰값: 표본크기 = 30

TEST1	TEST2	
	1	2
1	4	0
2	12	14

통계량:  $\chi^2 = 10.0833$       자유도:  $df = 1$   
 유의수준 ( $\alpha = 0.05$ ):  $\chi^2 = 3.84$       확률:  $P = 0.0015$

결론: 두 변수 간에 유의한 변화가 있습니다.

- ① TEST1 을 행, TEST2 를 열로 한 교차분할표를 보여준다.
- ②  $p = 0.0015 \Rightarrow$  두 변수 간에 통계적으로 유의한 비율의 차이(변화)가 있다.




## Exercise

### Cochran Q Test



진단법에 따라 진단결과에 차이가 있는지 알아보기 위하여 10 명의 환자에서 3 가지 진단법을 시행하였다. (1 = 양성, 2 = 음성) 진단법에 따라 결과가 다른지 검정하시오.

	Diagnostic Test		
	Test 1	Test 2	Test 3
1	1	1	1
2	1	1	0
3	0	1	0
4	1	1	0
5	0	0	0
6	1	1	0
7	1	1	1
8	1	1	0
9	0	0	1
10	0	1	0



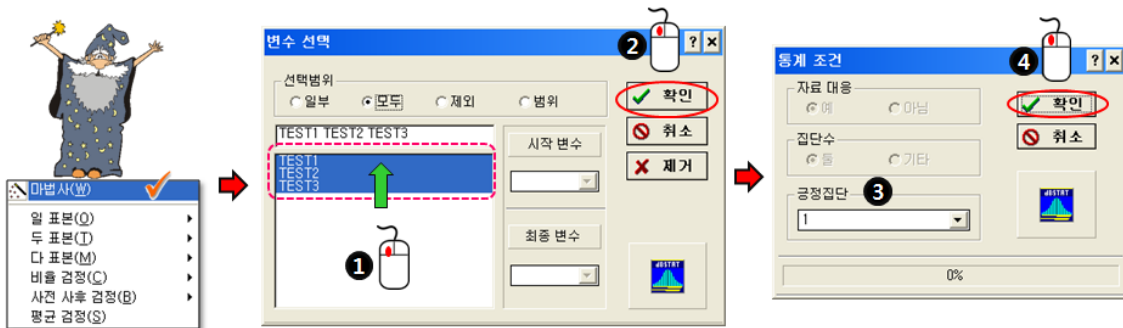
# dBSTAT

Data: Cochran.dbf

## 코크란 검정 (Cochran Q test)

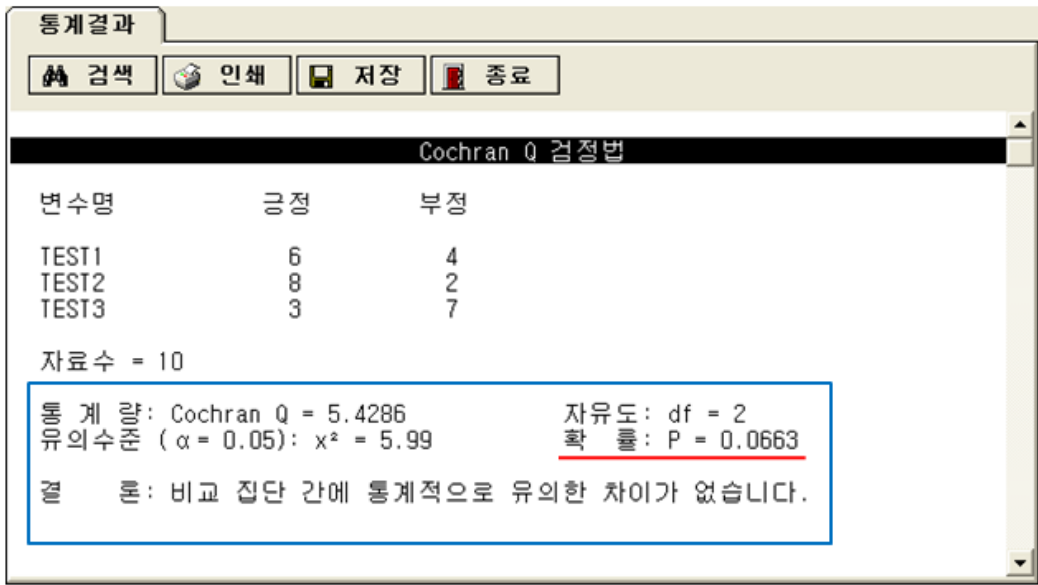
1 메뉴 → [통계] ▶ [비교통계] ▶ [마법사] 메뉴를 선택한다.

💡 [통계] ▶ [비교통계] ▶ [다 표본] ▶ [코크란 검정]을 선택해도 동일 결과가 출력된다.



- ① [변수 선택] 대화상자에서 변수 TEST1, TEST2, TEST3 을 선택한다.
- ② [확인] 버튼을 클릭하여 선택 작업을 마치면 [통계 조건] 대화상자가 나타난다.
- ③ [통계 조건] → [공정집단] ⇨ 1 항목을 선택한다. [확인] 버튼을 누른다.

## 2 통계결과



The image shows a screenshot of the SPSS '통계결과' (Statistical Results) window for a Cochran Q test. The window title is '통계결과' and it contains buttons for '검색' (Search), '인쇄' (Print), '저장' (Save), and '종료' (End). The main content area is titled 'Cochran Q 검정법' (Cochran Q Test). It displays a table with three columns: '변수명' (Variable Name), '긍정' (Positive), and '부정' (Negative). The data rows are TEST1, TEST2, and TEST3. Below the table, it shows '자료수 = 10'. A summary box contains the test statistics: '통계량: Cochran Q = 5.4286', '자유도: df = 2', '유의수준 (α = 0.05):  $\chi^2 = 5.99$ ', and '확률: P = 0.0663'. The conclusion states: '결론: 비교 집단 간에 통계적으로 유의한 차이가 없습니다.' (Conclusion: There is no statistically significant difference between the compared groups.)

변수명	긍정	부정
TEST1	6	4
TEST2	8	2
TEST3	3	7

자료수 = 10

통계량: Cochran Q = 5.4286      자유도: df = 2  
유의수준 (α = 0.05):  $\chi^2 = 5.99$       확률: P = 0.0663

결론: 비교 집단 간에 통계적으로 유의한 차이가 없습니다.

- ① 3 가지 진단법에 따른 긍정(1), 부정(0) 빈도수를 보여준다.
- ② Cochran Q = 5.429  $\Rightarrow$  p = 0.066 (>0.05) 진단법들 간에 통계적으로 유의한 차이가 없다.



## Exercise

### Chi-square test for trend

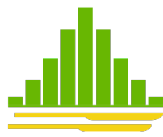
---



두 가지 치료방법에 대한 통증 치료효과에 대한 임상시험을 시행하였다.  
치료효과는 통증의 악화(worse), 동일(same), 호전(better)으로 나누어 조사하였다.  
치료방법과 치료효과 간에 연관성이 있는지 유의수준 5%에서 검정하시오.

**Table.** Clinical Trial of Two Treatments

치료	임상시험 결과			Total
	Worse	Same	Better	
Treatment A	10	45	80	135
Treatment B	2	15	35	52
Total	12	60	115	187

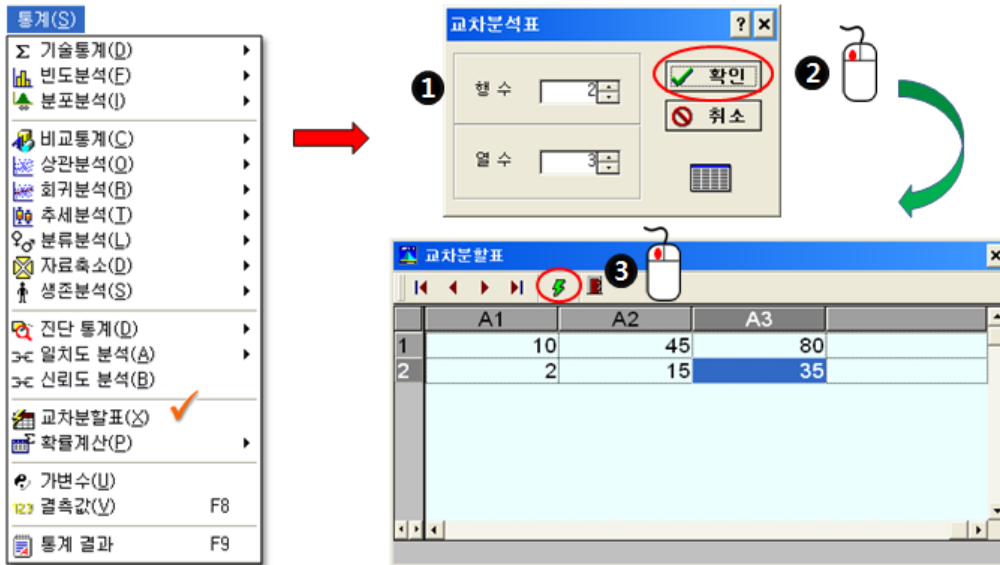


# dBSTAT

## Chi-square 추세검정

### 카이제곱 추세검정 (Chi-square test for trend)

1 메뉴 → [통계] ▶ [교차분할표] 메뉴를 선택하면 [교차분석표] 대화상자가 나타난다.



- ① [교차분석표] 대화상자에서 [행 수] ➡ 2, [열 수] ➡ 3 입력한다.
- ② [확인] 버튼을 클릭하여 선택 작업을 마치면 [교차분할표] 입력상자가 나타난다.
- ③ [교차분할표] 입력상자에 자료입력을 마치고 [확인] 버튼을 클릭한다.

## 2 통계결과

The screenshot shows a software window with tabs for '통계결과' (Statistical Results), '통계치' (Statistics), and '그래프' (Graphs). Below the tabs are buttons for '검색' (Search), '인쇄' (Print), '저장' (Save), and '종료' (Exit). The main content area is divided into two sections, each with a circled number in the left margin.

**1 [ Mantel-Haenszel Chi-Square Test for Trend ]**  
통계량:  $\chi^2 = 1.34$       자유도:  $df = 1$   
유의수준 ( $\alpha = 0.05$ ):  $\chi^2 = 3.84$       확률:  $P = 0.2471$   
결론: 두 변수 (열과 행) 간에 유의한 선형추세(linear trend)가 없습니다.  
주 의: 독립변수가 이항변수, 종속변수가 명목형 순위변수이어야 합니다.

**2 [ Cochran-Armitage Test for Trend ]**  
통계량:  $\chi^2 = 1.35$       자유도:  $df = 1$   
유의수준 ( $\alpha = 0.05$ ):  $\chi^2 = 3.84$       확률:  $P = 0.2459$   
결론: 두 변수 (열과 행) 간에 유의한 선형추세(linear trend)가 없습니다.  
주 의: 독립변수가 이항변수, 종속변수가 명목형 순위변수이어야 합니다.

① **Mantel-Haenszel** 추세검정:  $\chi^2 = 1.34$ ,  $P = 0.2471 > 0.05 \Rightarrow$  통계적 유의성이 없다.

② **Cochran-Armitage** 추세검정:  $\chi^2 = 1.35$ ,  $P = 0.2459 > 0.05 \Rightarrow$  통계적 유의성이 없다.







## Exercise

### Strength of Association



방사선 노출이 특정 질병과 관련성을 알아보려고 방사선 노출군 1000 명과 노출된 적이 없는 대조군 1000 명을 일정기간 관찰하여 질병의 발생여부를 조사하는 연구를 시행하였다. 방사선 노출과 특정 질병과의 연관성 정도를 구한다.

			Outcome	
			 present	 absent
Risk	 +	900	100	
	 -	200	800	



dBSTAT

Chi-square test

## 연관성 강도 (Strength of Association)

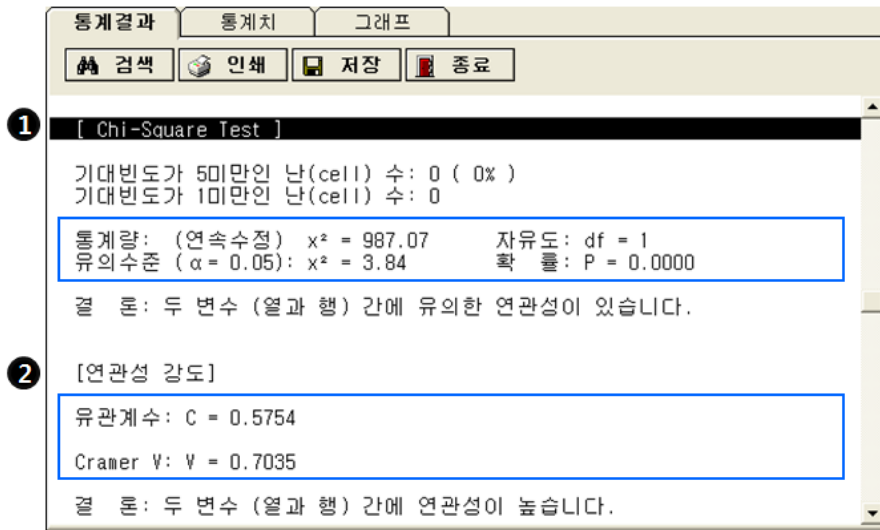
1 메뉴 → [통계] ▶ [교차분할표] 메뉴를 선택하면 [교차분석표] 대화상자가 나타난다.

The screenshot illustrates the steps to access the Chi-square test dialog box. On the left, the '통계(S)' menu is open, with '교차분할표(X)' selected. A red arrow points to the '교차분석표' dialog box, where '행 수' and '열 수' are both set to 2, and the '확인' button is circled in red. A green arrow points from the '확인' button to the '교차분할표' input window, which shows a 2x2 contingency table:

	A1	A2
1	900	100
2	200	800

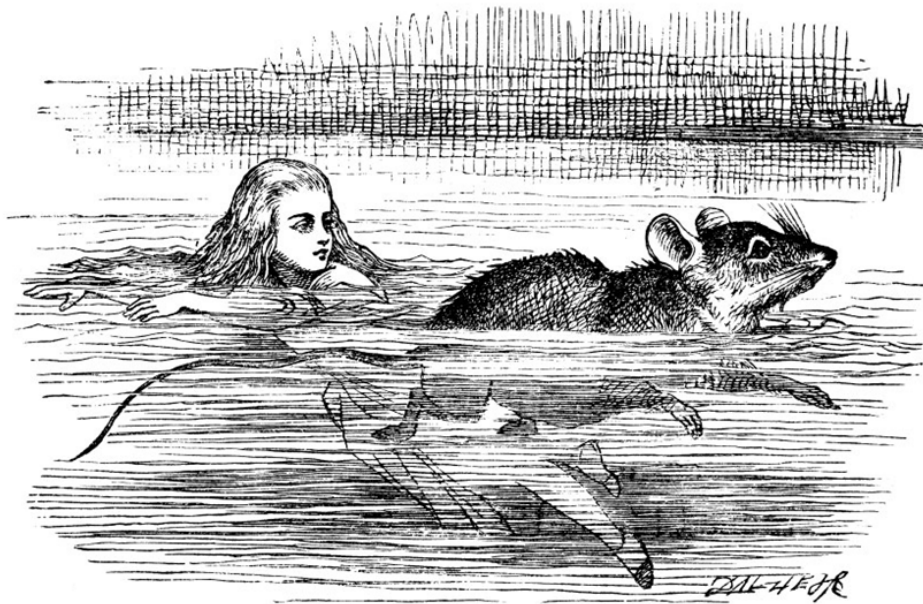
- ① [교차분석표] 대화상자에서 [행 수] ⇨ 2, [열 수] ⇨ 2 입력한다.
- ② [확인] 버튼을 클릭하여 선택 작업을 마치면 [교차분할표] 입력상자가 나타난다.
- ③ [교차분할표] 입력상자에 자료입력을 마치고 [미침] 버튼을 클릭한다.

## 2 통계결과



- ① **Chi-square** 검정:  $\chi^2 = 987.07$  (연속성수정),  $P < 0.0001 \Rightarrow$  통계적 유의성이 있다.
- ② **연관성 강도**: 분할계수  $C = 0.5754$ , Cramer  $V = 0.7035 > 0.7 \Rightarrow$  높은 연관성이 있다.

# Survival Analysis



*That will be a queer thing, to be sure!*

*However, everything is queer today.*



## Exercise

### Survival Analysis



심장병 환자들을 대상으로 생존시간(Time)과 생존 상태(Status)을 조사한 자료이다.

Kaplan-Meier 법에 의한 생존률과 생존 그래프를 그려보자.

(Time: 개월 단위, Status: 0 = 생존, 1 = 사망)

Patient	Time	Status	Patient	Time	Status
1	2	1	16	4	1
2	4	1	17	5	0
3	6	1	18	7	1
4	6	0	19	9	1
5	7	1	20	11	1
6	10	1	21	15	1
7	12	1	22	18	0
8	15	1	23	22	0
9	15	1	24	23	1
10	16	0	25	24	0
11	17	1	26	27	1
12	17	1	27	28	1
13	18	1	28	58	1
14	18	0	29	60	0
15	20	1	30	60	0



# dBSTAT

Data: survival.dbf

## 생존분석 (Survival Analysis)

1 메뉴 → [통계] ▶ [생존분석] ▶ [Kaplan-Meier] 메뉴를 선택한다.

[통계 조건] 대화상자가 나타난다.



- ① 시간 변수 ⇨ TIME 변수를 선택한다.
- ② 생존 변수 ⇨ STATUS, 종료집단 ⇨ 1 을 선택한다.
- ③ [확인] 버튼을 클릭하여 선택 작업을 마치면 [통계 결과]가 출력된다.

## 2 통계결과

통계결과		통계치		그래프		
🔍	검색	🔄	인쇄	💾	저장	
🛑	종료					
1 [ 생존통계 ]						
집 단	숫자	사망	중도절단	중도절단률 %		
Total	30	21	9	30.00		
2 [ 생존시간 ]						
집 단	평균 생존시간	95% 신뢰구간 하한값	상한값	중앙 생존시간	95% 신뢰구간 하한값	상한값
Total	23.466	15.638	31.293	17.000	11.950	22.050

① [생존통계] 전체 생존자료에 대한 사망과 중도절단 수를 보여준다.

② [생존시간] 평균과 중앙생존시간, 95% 신뢰구간을 보여준다.

## 생존률 (Survival Rate)

생존률 표 - Kaplan-Meier 법									
[ 생존률 ]									
집단 : 전체									
추적시점	① 생존	② 사망	③ 절단	사망률	생존률	누적생존률	표준오차	⑤ 95% 신뢰구간	
0	30	0	0	0.0000	1.0000	1.0000			
2	30	1	0	0.0333	0.9667	0.9667	0.0328	0.786, 0.995	
4	29	2	0	0.0690	0.9310	0.9000	0.0548	0.721, 0.967	
5	27	0	1						
6	26	1	1	0.0385	0.9615	0.8654	0.0627	0.680, 0.947	
7	24	2	0	0.0833	0.9167	0.7933	0.0754	0.596, 0.902	
9	22	1	0	0.0455	0.9545	0.7572	0.0801	0.556, 0.876	
10	21	1	0	0.0476	0.9524	0.7212	0.0840	0.518, 0.850	
11	20	1	0	0.0500	0.9500	0.6851	0.0872	0.481, 0.822	
12	19	1	0	0.0526	0.9474	0.6490	0.0898	0.445, 0.794	
15	18	3	0	0.1667	0.8333	0.5409	0.0941	0.343, 0.703	
16	15	0	1						
17	14	2	0	0.1429	0.8571	0.4636	0.0952	0.273, 0.634	
18	12	1	2	0.0833	0.9167	0.4250	0.0948	0.240, 0.598	
20	9	1	0	0.1111	0.8889	0.3777	0.0953	0.198, 0.557	
22	8	0	1						
23	7	1	0	0.1429	0.8571	0.3238	0.0957	0.152, 0.509	
24	6	0	1						
27	5	1	0	0.2000	0.8000	0.2590	0.0960	0.099, 0.454	
28	4	1	0	0.2500	0.7500	0.1943	0.0913	0.056, 0.393	
58	3	1	0	0.3333	0.6667	0.1295	0.0806	0.024, 0.325	
60	2	0	2	0.0000	1.0000	0.1295	0.0806	0.024, 0.325	

- ① **Number at risk:** 추적시점 시작 시 생존자, 즉 사망 위험에 노출된 대상자 수
- ② **Uncensored data:** 추적시점에서 발생한 사망자 수, 즉 종결된 자료 수이다.
- ③ **Censored data:** 추적시점에서 중도 절단된 자료 수를 뜻한다.
- ④ **누적생존률:** 추적시점에서 계산된 생존률로 4 개월 누적생존률은 90%이다.
- ⑤ 누적생존률에 대한 95% 신뢰구간이다. 4 개월 95% CI = (0.72, 0.97)

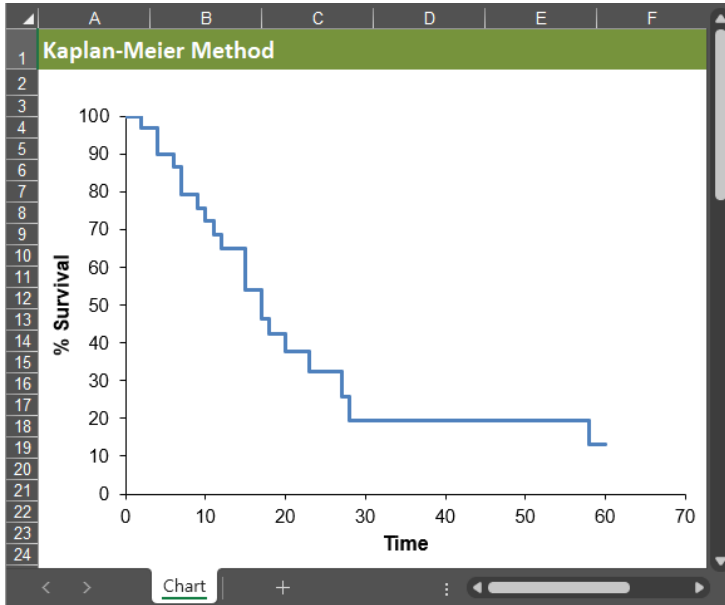


추적시점 시작 시 생존자료 수(**number at risk**) < 10 이면 누적생존률을 신뢰할 수 없으므로 10 이상인 시점까지만 Kaplan-Meier 분석을 하여야 한다.

위의 통계결과 생존률 표에서 ➡ 표시된 추적시점(18 개월)까지만 분석하는 것이 올바른 통계분석 방법이다.

### 3 Excel Chart

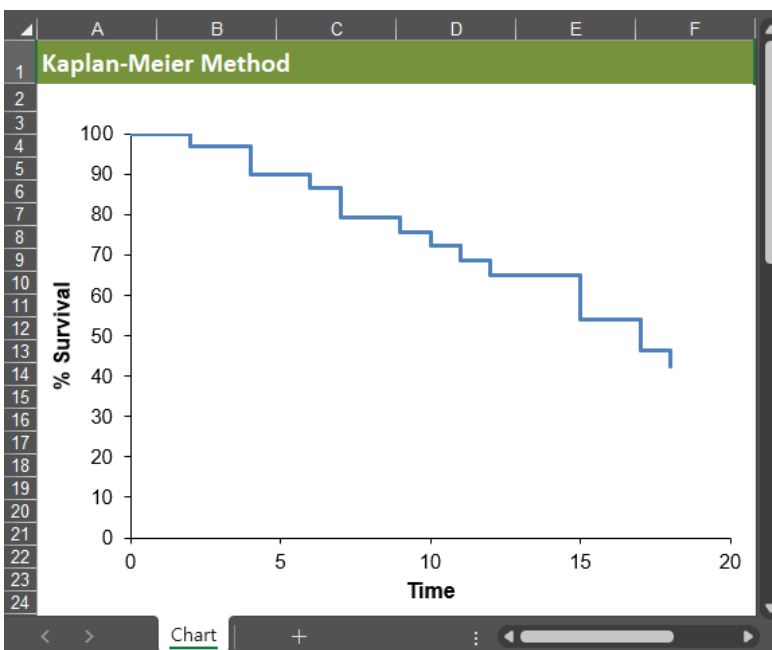
Kaplan-Meier 생존 그래프(survival curve)가 나타난다.



✎ 생존 그래프에서 60 개월 추적시점에서 생존자료수가 2 로 신뢰할 수 없으므로 생존자료 수  $\geq 10$  시점 (18 개월) 까지만 그래프를 작성하여야 한다.

[통계 조건] 대화상자에서 시간 변수  $\rightarrow$  STATUS, 종료  $\rightarrow$  18 입력하면 생존 그래프를 추적시점 18 개월까지만 그린다.

다음은 추적기간(Time) 18 개월인 Kaplan-Meier 생존 그래프이다.





## Exercise

### Survival Comparison

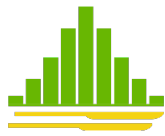
---



심장병 환자들을 대상으로 기존 치료법과 새로운 치료법(Tx)을 시행한 후에 생존시간(Time)과 생존상태(Status)을 조사한 자료이다. 두 가지 치료법에 따른 생존률을 비교하여 보자.

Patient	Tx	Time	Status	Patient	Tx	Time	Status
1	1	2	1	16	2	4	1
2	1	4	1	17	2	5	0
3	1	6	1	18	2	7	1
4	1	6	0	19	2	9	1
5	1	7	1	20	2	11	1
6	1	10	1	21	2	15	1
7	1	12	1	22	2	18	0
8	1	15	1	23	2	22	0
9	1	15	1	24	2	23	1
10	1	16	0	25	2	24	0
11	1	17	1	26	2	27	1
12	1	17	1	27	2	28	1
13	1	18	1	28	2	58	1
14	1	18	0	29	2	60	0
15	1	20	1	30	2	60	0

Tx: 1 = 기존 치료, 2 = 새 치료, Time: 개월, Status: 0 = 생존, 1 = 사망



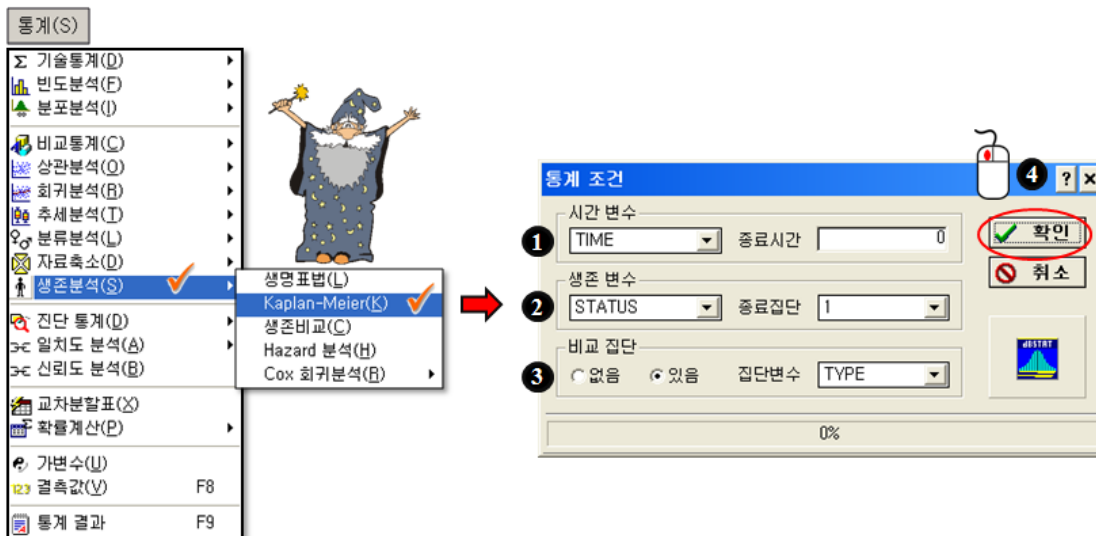
# dBSTAT

Data: survival.dbf

## 생존률 비교

1 메뉴 → [통계] ▶ [생존분석] ▶ [Kaplan-Meier] 메뉴를 선택한다.

[통계 조건] 대화상자가 나타난다.



- ① 시간 변수 ⇨ TIME 변수를 선택한다.
- ② 생존 변수 ⇨ STATUS, 종료집단 ⇨ 1 을 선택한다.
- ③ 비교 집단 ⇨ TYPE 변수를 선택한다.
- ④ [확인] 버튼을 클릭하여 선택 작업을 마치면 [통계 결과]가 출력된다.

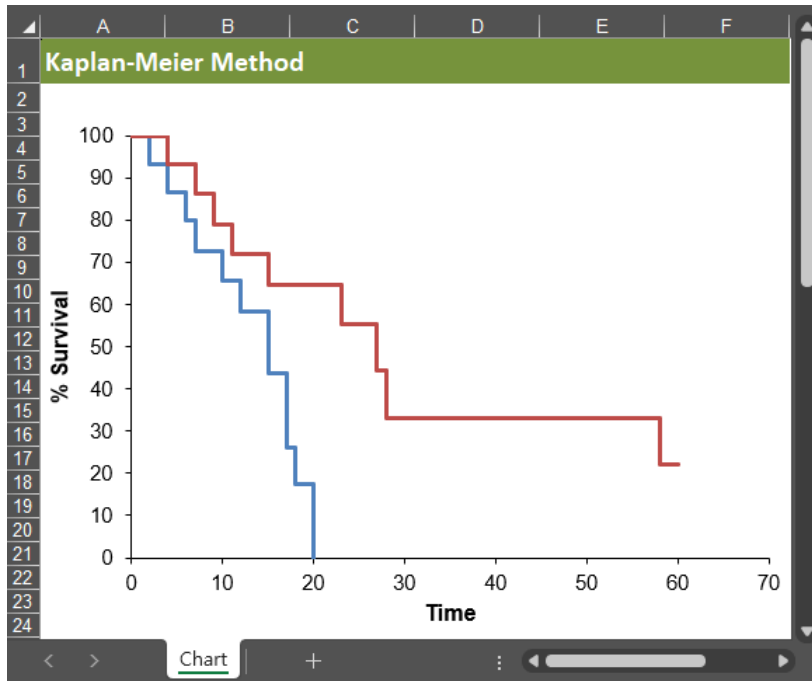
## 2 통계결과

통계결과				
🔍 검색	📄 인쇄	💾 저장	🛑 종료	
[ 생존 통계 ]				
변 수 명: TYPE				
집 단	숫자	사망	중도절단	중도절단률 %
1	15	12	3	20.00
2	15	9	6	40.00
Total	30	21	9	30.00
[ 생존 시 간 ]				
집 단	평균 생존시간	95% 신뢰구간 하한값 상한값	중앙 생존시간	95% 신뢰구간 하한값 상한값
TYPE				
1	13.120	9.970 16.270	15.000	9.598 20.402
2	31.214	19.128 43.301	27.000	16.330 37.670
Log-Rank (Mantel-Cox) 검 정 법				
통계량: $\chi^2 = 6.179$		자유도: $df = 1$		
유의수준 ( $\alpha = 0.05$ ): $\chi^2 = 3.841$		확 률: $P = 0.0129$		
Yates 연속성수정: Chi-square = 4.97				
결 론: 집단들간의 생존분포가 동일하지 않습니다.				

- ① Type = 1, Type = 2 두 집단에서 생존시간의 평균과 95% 신뢰구간을 보여준다.
- ② 생존시간의 기술통계는 중앙값을 사용하여야 한다.
- ③ Log rank 검정결과 Log rank test:  $\chi^2 = 6.18$ ,  $p = 0.013$  로 통계적으로 유의하다.

### 3 Excel Chart

Type = 1, Type = 2 두 집단에서 누적생존률을 보여준다.



### 생존표

다음은 Type = 1 (기존 치료법)에 대한 생존표이다.

통계결과

🔍 검색   📄 인쇄   💾 저장   🛑 종료

생존 통계 - Kaplan-Meier 법

[ 생존률 ]

집단1 : TYPE = 1

추적시점	생존	사망	절단	사망률	생존률	누적생존률	표준오차	95% 신뢰구간
0	15	0	0	0.0000	1.0000	1.0000		
2	15	1	0	0.0667	0.9333	0.9333	0.0644	0.613, 0.990
4	14	1	0	0.0714	0.9286	0.8667	0.0878	0.564, 0.965
6	13	1	1	0.0769	0.9231	0.8000	0.1033	0.500, 0.931
7	11	1	0	0.0909	0.9091	0.7273	0.1167	0.425, 0.888
10	10	1	0	0.1000	0.9000	0.6545	0.1257	0.357, 0.840
12	9	1	0	0.1111	0.8889	0.5818	0.1311	0.294, 0.787
15	8	2	0	0.2500	0.7500	0.4364	0.1327	0.182, 0.667
16	6	0	1					
17	5	2	0	0.4000	0.6000	0.2618	0.1244	0.068, 0.512
18	3	1	1	0.3333	0.6667	0.1745	0.1093	0.029, 0.422
20	1	1	0	1.0000	0.0000			

① 생존 ▶ 추적시점 시작 시 생존 수(number at risk)이다.

② 추적시점 ➡ 10 개월 생존 수 = 10 이다.



추적시점 생존 수 10 이상인 시점인 10 개월까지만 분석하는 것이 올바른 방법이다.

## 생존 종료시간

The image shows a software dialog box titled "통계 조건" (Statistical Conditions). It has a standard window header with a question mark and a close button. The dialog is organized into several sections:

- 시간 변수** (Time Variable): A dropdown menu set to "TIME". To its right is a text input field labeled "종료시간" (End Time) containing the number "10". This field is enclosed in a red dashed rectangular box.
- 생존 변수** (Survival Variable): A dropdown menu set to "STATUS". To its right is another dropdown menu labeled "종료집단" (End Group) set to "1".
- 비교 집단** (Comparison Group): Two radio buttons, "없음" (None) and "있음" (Yes), with "있음" selected. To the right is a dropdown menu labeled "집단변수" (Group Variable) set to "TYPE".
- Buttons:** On the right side, there are two buttons: "확인" (OK) with a green checkmark icon, and "취소" (Cancel) with a red prohibition sign icon.
- Progress Bar:** At the bottom of the dialog, there is a progress bar showing "0%".

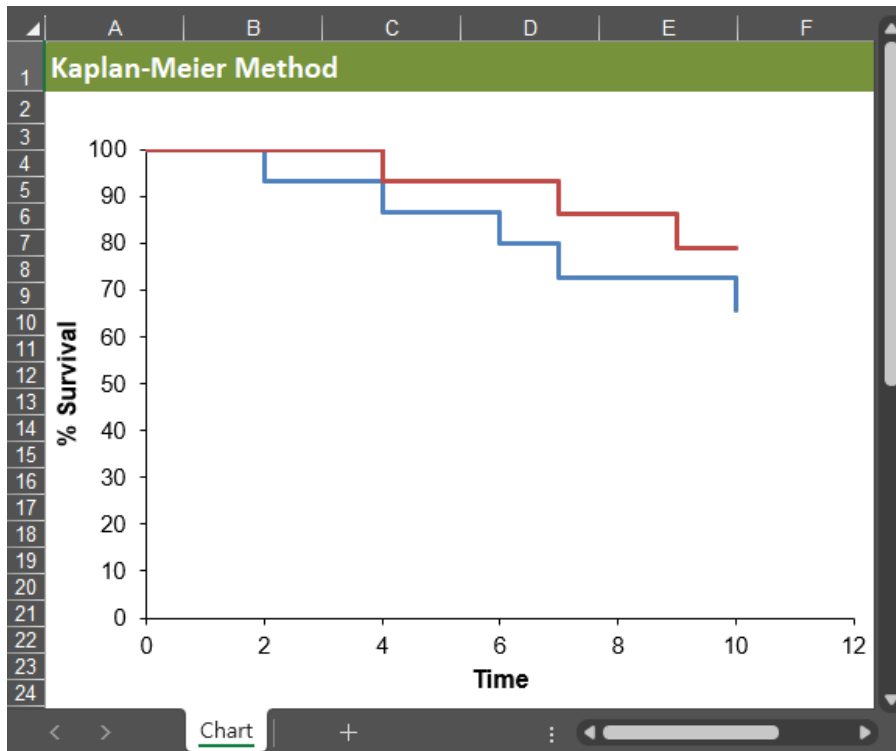
① 시간 변수 ➡ TIME ▶ 종료시간 ➡ 10 을 입력한다.

② 생존 변수 ➡ STATUS, 종료집단 ➡ 1 을 선택한다.

③ 비교 집단 ➡ TYPE 변수를 선택한다.

④ [확인] 버튼을 클릭하여 선택 작업을 마치면 [통계 결과]가 출력된다.

## 생존 종료시점



생존분석 종료시점을 10 개월로 정하여 재분석한 결과 Log rank test  $\chi^2 = 0.24$ ,  $p = 0.396$

이므로 통계적으로 유의하지 않다.





## Exercise

### Cox Regression



심장병 환자들을 대상으로 기존 치료법과 새로운 치료법(Tx)을 시행한 후에 생존시간(Time), 생존상태(Status), 연령(Age)을 조사한 자료이다.  
치료법과 연령이 심장병 환자의 사망위험과 연관이 있는지 알아보자.

Patient	Tx	Age	Time	Status	Patient	Tx	Age	Time	Status
1	1	68	4	1	16	2	63	2	1
2	1	55	5	0	17	2	60	4	1
3	1	67	7	1	18	2	66	6	1
4	1	56	9	1	19	2	59	6	0
5	1	70	11	1	20	2	68	7	1
6	1	62	15	1	21	2	54	10	1
7	1	53	18	0	22	2	51	12	1
8	1	48	22	0	23	2	69	15	1
9	1	59	23	1	24	2	50	15	1
10	1	51	24	0	25	2	62	16	0
11	1	60	27	1	26	2	58	17	1
12	1	59	28	1	27	2	54	17	1
13	1	69	58	1	28	2	52	18	1
14	1	62	60	0	29	2	48	18	0
15	1	51	60	0	30	2	59	20	1

Tx: 1 = 기존치료, 2 = 새 치료, Time: 개월, Status: 0 = 생존, 1 = 사망



**dBSTAT**

Data: survival.dbf

## Cox 회귀분석 (Cox Regression)

1 메뉴 → [통계] ▶ [생존분석] ▶ [Cox 회귀분석] ▶ [회귀분석] 메뉴를 선택한다.

[통계 조건] 대화상자가 나타난다.



- ① 시간 변수 ⇒ TIME 변수를 선택한다.
- ② 생존 변수 ⇒ STATUS, 종료집단 ⇒ 1 을 선택한다.
- ③ 독립 변수 ⇒ AGE 변수를 선택한다.
- ④ 집단 변수 ⇒ TYPE 변수를 선택한다.
- ⑤ [확인] 버튼을 클릭하여 선택 작업을 마치면 [통계 결과]가 출력된다.

## 2 통계결과

[통계결과] 창에 Cox 회귀분석 검정 결과를 보여준다.

**1** [Cox 회귀 모형의 적합성]

-2 Log Likelihood : 101.57

Chi-square	자유도	확률
11.4236	2	0.0033

결론: Cox 회귀분석 모형이 적합합니다.

**2** [Cox 회귀 방정식]

LOG h(t)/ho(t) = B1 x X1 + B2 x X2 + ... + Bn x Xn (Bn: 회귀계수, Xn: 공변량)

공변량(X)	회귀계수(B)	표준오차	Wald	유의성	상관계수	위험비(HR)	95% 신뢰구간
AGE	0.0804	0.0347	5.3793	0.0204*	0.1729	1.0837	1.0125, 1.1598
TYPE(1)	-1.4314	0.5567	6.6096	0.0101*	-0.2020	0.2390	0.0803, 0.7117

**3** 결론: + 표시(P < 0.05) 된 공변량이 통계적으로 의미가 있습니다.

[Categorical Variables]

TYPE	TYPE(0)	TYPE = 1
	TYPE(1)	TYPE = 2

### ① [Cox 회귀 모형의 적합성]

로그우도비  $-2LL = 101.57 \Rightarrow \chi^2 = 11.42, p = 0.003$  으로 통계적으로 유의하다.

### ② [Cox 회귀 방정식]

회귀계수에 대한 Wald 검정, 유의확률, 위험비와 95% 신뢰구간을 보여준다.

TYPE(1)의 위험비 = 0.239 의 의미는 Type=2 의 Type=1 에 대한 비교위험을 뜻한다.

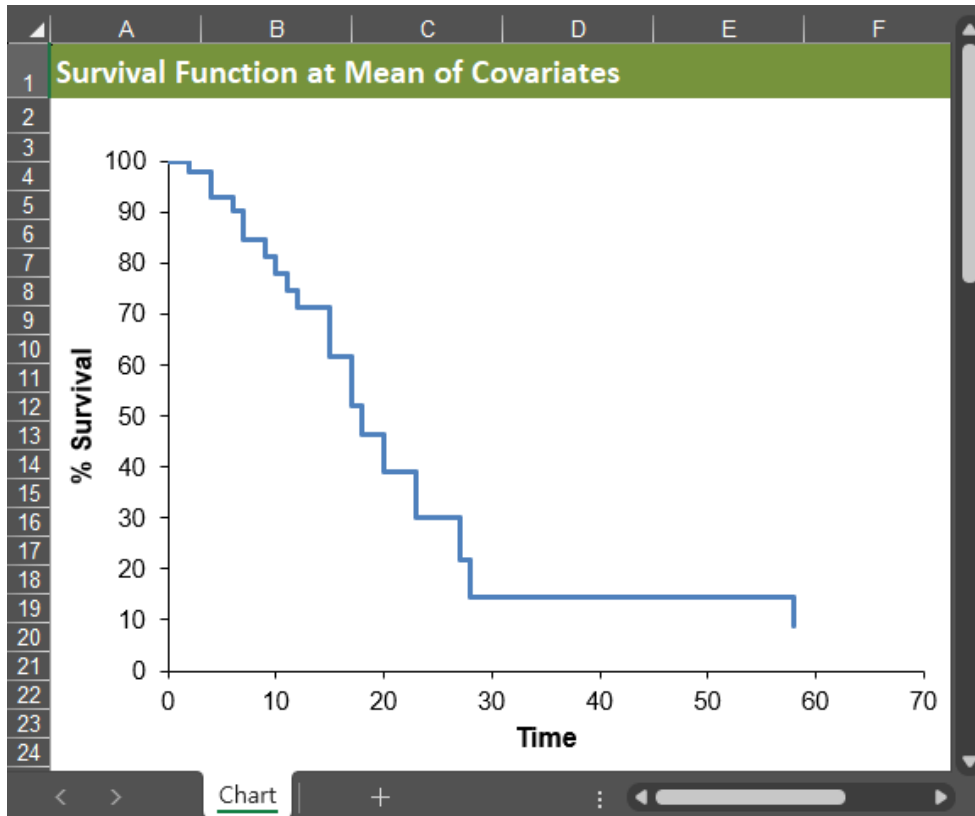
### ③ [결론]

Age, Type 공변량에 대한 Wald 검정 결과  $p < 0.05$  이므로 통계적으로 유의하다.

**[Categorical Variables]:** 범주형 명목변수를 가변수화하여 기준범주 TYPE(0): TYPE=1  $\Rightarrow$  0, TYPE(1): TYPE=2  $\Rightarrow$  1 로 바뀌었음을 보여준다. 회귀식에 입력할 때는 가변수값을 사용한다.

### 3 Excel Chart

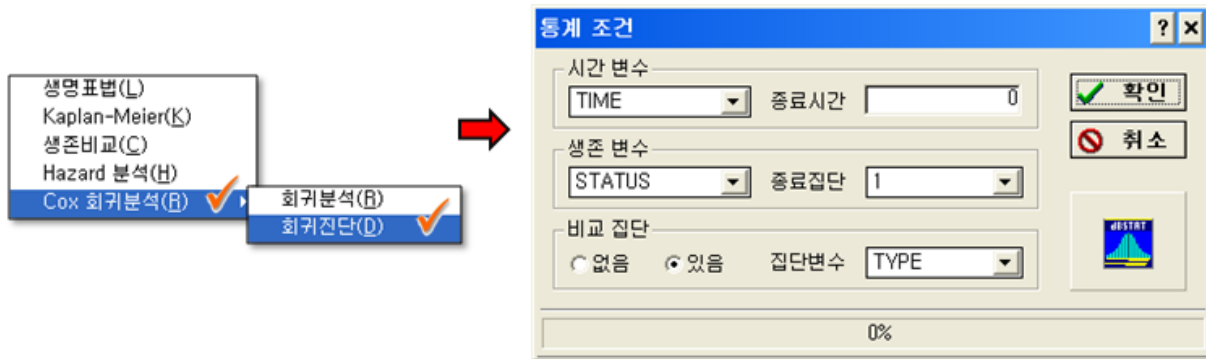
공변량들의 평균에서의 생존곡선을 보여준다.



## 4 Cox 회귀분석 가정

[통계] ▶ [생존분석] ▶ [Cox 회귀분석] ▶ [회귀진단] 메뉴를 선택한다.

[통계 조건] 대화상자가 나타난다.



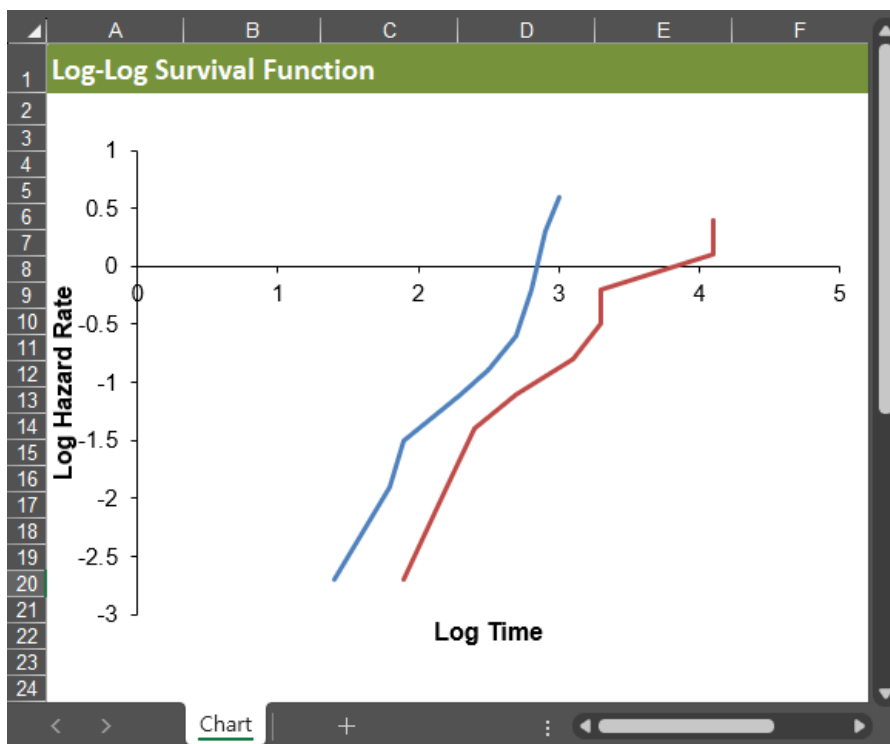
[Kaplan-Meier] 분석과 동일한 대화상자가 나타난다.

[시간 변수], [생존 변수], [비교집단] 상자에 변수를 입력하고 [확인] 버튼을 누른다.

### 로그-로그 생존곡선

X 축은  $\log(\text{time})$ , Y 축은  $\log[-\log(S(t))]$  (log hazard)인 log-log 그래프를 보여준다.

Type 에 따른 두 그래프가 평행선에 가까우므로 비례위험 가정을 만족한다고 할 수 있다.



# Sample Size



*“Five and Seven said nothing, but looked at Two.”*



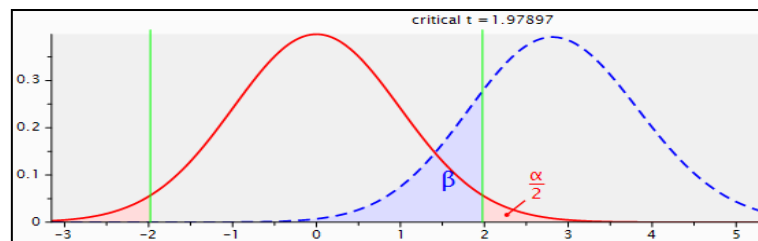
## Exercise

### Sample Size



비만 환자에서 새로운 Diet 법이 기존 방법에 비하여 체중 감소 효과가 있는지 알아보는 임상시험을 계획하였다. 1년간 시행 후 시험군과 대조군의 평균 체중 차이가 최소 2 kg 이면 치료효과가 있다고 판정한다. 선행연구에서 기존 방법의 평균 체중감소는 8 kg, 표준편차는 4 kg 으로 알려져 있다.

$\alpha = 0.05$ , 검정력 80%로 설정하였을 때 필요한 연구대상자 수를 구하시오.





## dBSTAT

### 표본크기 (Sample Size)

1 메뉴 → [통계] ▶ [표본크기] ▶ [평균 비교] ▶ [독립 표본] 메뉴를 선택한다.

- ⑥ [검정] ▶ 단측, 양측 선택 [검정력] ▶ alpha ⇒ 0.05, 검정력 ⇒  $1 - \text{beta} = 1 - 0.2 = 0.8$  (80%)
- ⑦ [효과크기] 평균차이 ⇒ 2, 표준편차 ⇒ 4 입력한다.
- ⑧ [집단비율] 시험군:대조군 ⇒ 1:1, 탈락률 ⇒ 0 입력한다.
- ⑨ [Study Design] ⊙우위성, ○동등성, ○비열등성 ⇒ 우위성 선택. 허용한계 ⇒ 0 입력한다.
- ⑩ [확인] 버튼을 클릭하면 통계결과 창이 나타난다.

## 2 통계결과

The screenshot shows the '통계결과' (Statistical Results) window in SPSS. The title bar reads 'Sample Size for Two Sample Comparison of Means'. The window contains two sections:

**1 [Independent t-test]**

연구설계	평균차이	표준편차	허용한계	유의수준	검정력
Equality	2.00	4.00	0.00	0.05	0.80

**2 [Sample Size]**

표본 크기	시험군	대조군	탈락률
126	63	63	0.00

③ [Independent t-test] Equality test, 유의수준 5%, 검정력 80%

④ [Sample Size] 시험군 63 명, 대조군 63 명으로 연구대상자 수는 총 126 명이다.



## Exercise

### Sample Size

---



난치성 질환 환자들을 대상으로 치료약이 합병증을 낮추는 효과가 있는지  
알아보는 임상시험을 계획하였다. 선행연구에서 합병증은 40%로 알려져 있다.  
치료효과는 합병증을 10% 감소시키면 임상적으로 유의하다고 판정한다. 유의수준  
 $\alpha = 0.05$ , 검정력 80%로 설정하였을 때 필요한 연구대상자 수를 구하시오.



dBSTAT

## 표본크기

1 메뉴 → [통계] ▶ [표본크기] ▶ [비율 비교] 메뉴를 선택한다.

Sample Size

1  단측  양측    검정력 alpha  beta     5  확인

2 효과크기

대조군 %  시험군 %      취소

3 집단비율

시험군 : 대조군  :     탈락률

4 Study Design

우위성  동등성  비열등성    허용한계 %

- ① 검정 > 단측, 양측 선택 [검정력] > alpha → 0.05, 검정력 →  $1 - \text{beta. } (0.2) = 0.8 (80\%)$
- ② [효과크기] 대조군 % → 30, 시험군 % → 40 입력한다.
- ③ [집단비율] 시험군:대조군 → 1:1, 탈락률 → 0 입력한다.
- ④ [Study Design] 우위성, 동등성, 비열등성 → 우위성 선택. 허용한계 % → 0 입력한다.
- ⑤ [확인] 버튼을 클릭하면 통계결과 창이 나타난다.

## 2 통계결과

Sample Size for Two Sample Comparison of Proportions					
[z-test]					
연구설계	대조군 비율	시험군 비율	허용한계	유의수준	검정력
Equality	30.00	40.00	0.00	0.05	0.80
[Sample Size]					
표본 크기	시험군	대조군	탈락률		
712	356	356	0.00		

① [z-test] Equality test, 유의수준 5%, 검정력 80%

② [Sample Size] 시험군 356 명, 대조군 356 명으로 연구대상자 수는 총 712 명이다