8th Week

# Two-Sample Hypothesis Test: Comparing Two Means

# 두 평균의 비교: 독립표본과 대응표본

2011

**Jongseok Lee**

**Business Administration**

**Hallym University**

# 변수 형태와 통계적 분석방법

$H_0 : X \perp Y$

$H_1 : X \sim Y$

| X | Y | |
|---|---|---|
| Categorical | Categorical | **Chi-square Test** |
| Categorical | Numerical | **집단들 사이의 평균 비교** |
| Numerical | Numerical | |

두 개의 집단 혹은 범주

☞ 독립표본 / 대응표본 T 검정

두 개 이상의 집단 혹은 범주

☞ ANOVA

골다공증과 같은 골밀도와 관련된 질병에
남성보다 여성이 취약한가?

**X**
Male *or* Female

~

**Y**
L-spine BMD

**Categorical** **Numerical**

| Variables | Male | Female | *t*-value | *p*-value |
|---|---|---|---|---|
| | *n* = 547 | *n* = 2,106 | | |
| L-spine BMD | 1.160 ± 0.153 | 1.062 ± 0.174 | 12.963 | 0.000 |

사람들의 요추 골밀도와 대퇴 골밀도는
그 평균에 차이가 있는가?

**X**
L-spine *or* Femur

~

**Y**
Bone Mineral Density

**Categorical** **Numerical**

| Variables | L-spine BMD | Femur BMD | *t*-value | *p*-value |
|---|---|---|---|---|
| | *n* = 2,461 | *n* = 2,461 | | |
| Total | 1.082 ± 0.175 | 0.944 ± 0.137 | 56.702 | 0.000 |

# Act 1

골다공증과 같은 골밀도와 관련된 질병에
남성보다 여성이 취약한가?

**X**
Male *or* Female

~

**Y**
L-spine BMD

**Categorical** **Numerical**

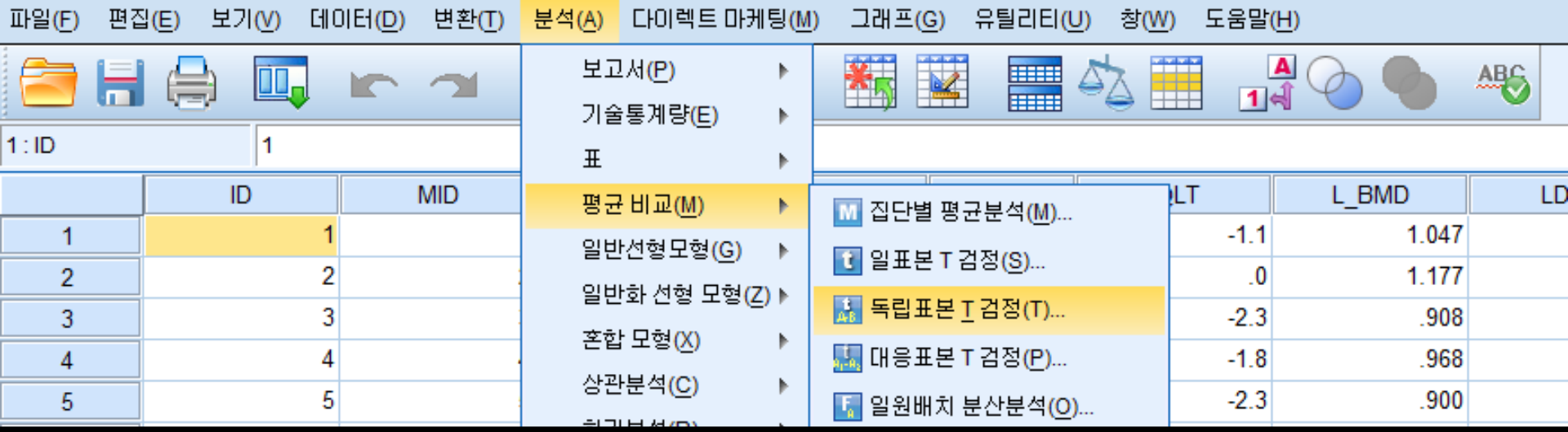
| Variables | Male | Female | *t*-value | *p*-value |
|---|---|---|---|---|
| | *n* = 547 | *n* = 2,106 | | |
| L-spine BMD | 1.160 ± 0.153 | 1.062 ± 0.174 | 12.963 | 0.000 |

# 독립표본 T 검정

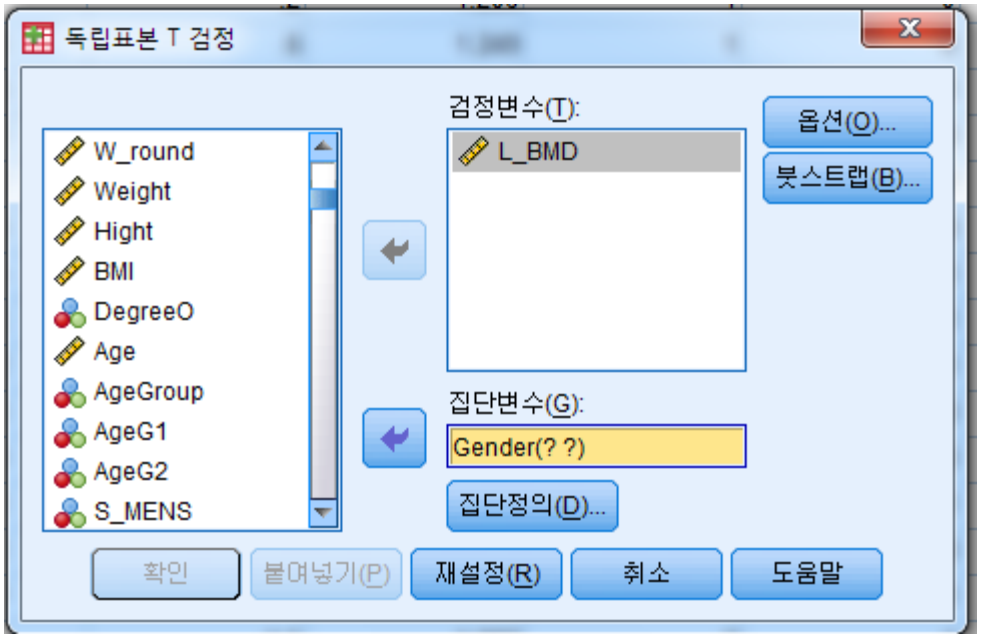
## 자료의 구조

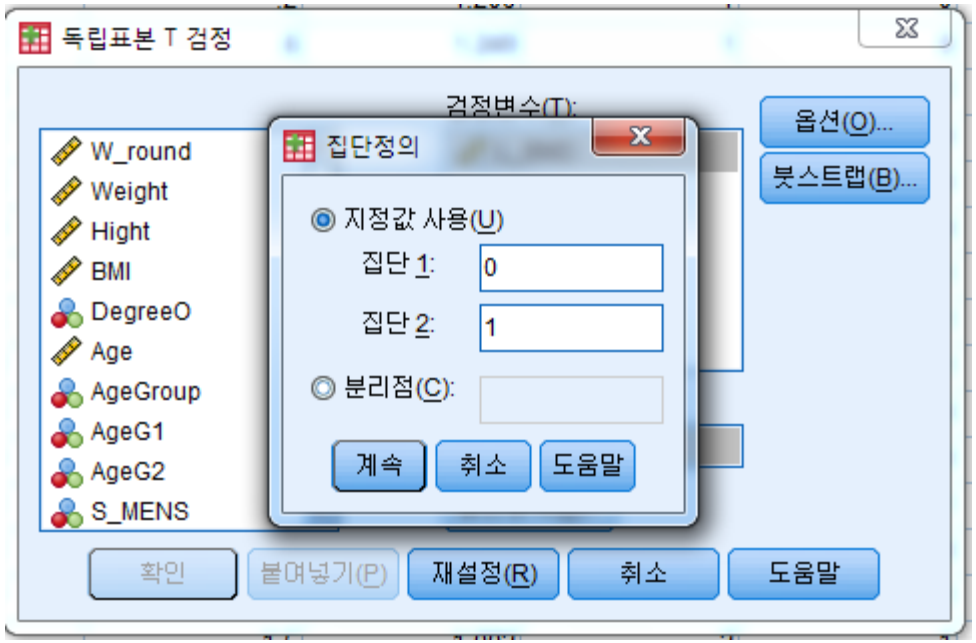
| Age | Gender |
|---|---|
| 31 | 0 |
| 30 | 0 |
| 30 | 0 |
| 30 | 0 |
| 30 | 0 |
| 30 | 0 |
| 29 | 0 |
| 29 | 0 |
| 29 | 0 |
| 28 | 0 |
| 28 | 0 |
| 26 | 0 |
| 64 | 1 |
| 56 | 1 |
| 43 | 1 |
| 57 | 1 |
| 64 | 1 |
| 50 | 1 |
| 54 | 1 |
| 48 | 1 |
| 48 | 1 |
| 47 | 1 |
| 44 | 1 |
| 53 | 1 |
| 68 | 1 |
| 42 | 1 |
| 48 | 1 |

두 집단의 평균 비교




# SPSS를 이용한 독립표본 T 검정




독립표본 T 검정
검정변수(T):
W_round
Weight
Hight
BMI
DegreeO
Age
AgeGroup
AgeG1
AgeG2
S_MENS
옵션(O)...
붓스트랩(B)...
집단정의
지정값 사용(U)
집단 1:
0
집단 2:
1
분리점(C):
계속
취소
도움말
확인
붙여넣기(P)
재설정(R)
취소
도움말

# SPSS 분석 결과

**집단통계량**

| | Gender | N | 평균 | 표준편차 | 평균의 표준오차 |
|---|---|---|---|---|---|
| L_BMD | 0 | 547 | 1.16036 | .153393 | .006559 |
| | 1 | 2106 | 1.06210 | .174396 | .003800 |

**독립표본 검정**

| | | Levene의 등분산 검정 | | 평균의 동일성에 대한 t-검정 | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | 차이의 95% 신뢰구간 | |
| | | F | 유의확률 | t | 자유도 | 유의확률 (양쪽) | 평균차 | 차이의 표준오차 | 하한 | 상한 |
| L_BMD | 등분산이 가정됨 | 14.752 | .000 | 12.024 | 2651 | .000 | .098260 | .008172 | .082236 | .114284 |
| | 등분산이 가정되지 않음 | | | 12.963 | 946.489 | .000 | .098260 | .007580 | .083384 | .113136 |

**먼저 남성과 여성, 두 집단의 분산이 같은지 알아보기 위해 F 검정 결과를 보아야 한다. 그 결과 $p$-값이 0.000이 되어 두 집단의 분산이 같다는 귀무가설을 유의수준 0.001에서 기각할 수 있다. 따라서 "등분산이 가정되지 않음"의 결과를 보아야 한다.**

# 보고서의 작성

**Table. Mean Comparisons of L-BMD by Gender**

| Variables | Male | Female | *t*-value | *p*-value |
|---|---|---|---|---|
| | *n* = 547 | *n* = 2,106 | | |
| L-spine BMD | 1.160 ± 0.153 | 1.062 ± 0.174 | 12.963 | 0.000 |

**남성과 여성 집단에 요추골밀도 평균의 차이가 있는지 알아보기 위하여 독립표본 T 검정을 수행하였다. 남성 집단의 요추골밀도 평균은 1.160이고, 여성 집단의 요추골밀도 평균은 1.062인 것으로 나타났다. 그리고 이러한 차이는 유의수준 0.001에서 통계적으로 유의한 것이었다.**

**집단통계량**

| | Gender | N | 평균 | 표준편차 | 평균의 표준오차 |
|---|---|---|---|---|---|
| L_BMD | 0 | 547 | 1.16036 | .153393 | .006559 |
| | 1 | 2106 | 1.06210 | .174396 | .003800 |

**독립표본 검정**

| | | Levene의 등분산 검정 | | 평균의 동일성에 대한 t-검정 | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | 차이의 95% 신뢰구간 | |
| | | F | 유의확률 | t | 자유도 | 유의확률 (양쪽) | 평균차 | 차이의 표준오차 | 하한 | 상한 |
| L_BMD | 등분산이 가정됨 | 14.752 | .000 | 12.024 | 2651 | .000 | .098260 | .008172 | .082236 | .114284 |
| | 등분산이 가정되지 않음 | | | 12.963 | 946.489 | .000 | .098260 | .007580 | .083384 | .113136 |

# Act 2

사람들의 요추 골밀도와 대퇴 골밀도는
그 평균에 차이가 있는가?

**X**
L-spine *or* Femur

~

**Y**
Bone Mineral Density

**Categorical** **Numerical**

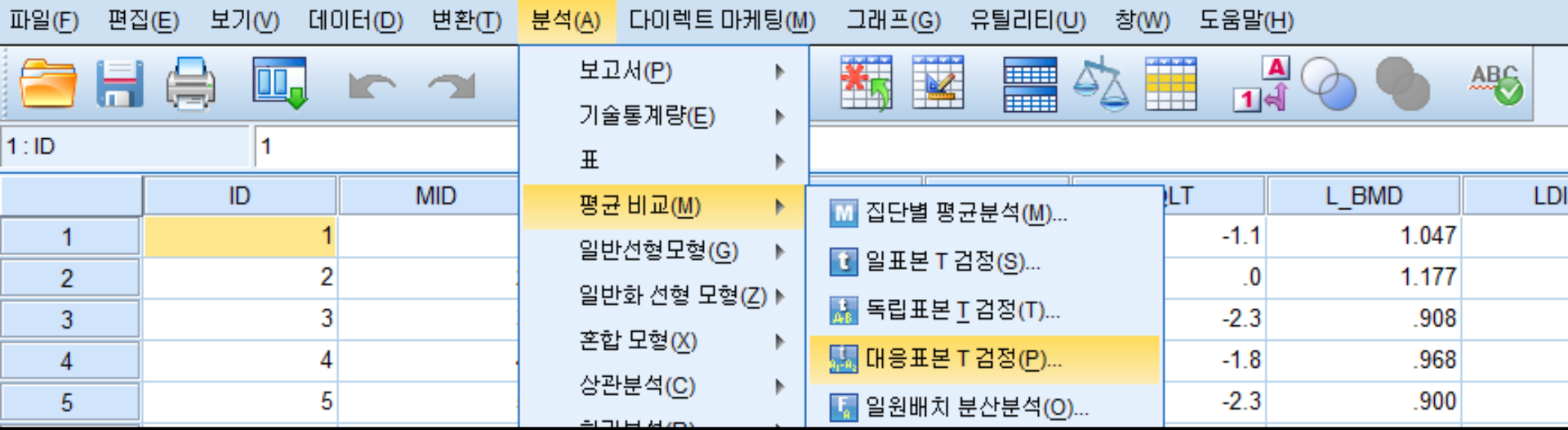
| Variables | L-spine BMD | Femur BMD | *t*-value | *p*-value |
|---|---|---|---|---|
| | *n* = 2,461 | *n* = 2,461 | | |
| Total | 1.082 ± 0.175 | 0.944 ± 0.137 | 56.702 | 0.000 |

# 대응표본 T 검정

## 자료의 구조

| L_BMD | F_BMD |
|---|---|
| 1.047 | .865 |
| 1.177 | .952 |
| .908 | .822 |
| .968 | .817 |
| .900 | .834 |
| 1.206 | .910 |
| 1.249 | .833 |
| .905 | .925 |
| 1.100 | .836 |
| 1.311 | .899 |
| 1.157 | .951 |
| .990 | .807 |
| .812 | .752 |
| .949 | 1.058 |
| 1.226 | .899 |
| 1.188 | 1.150 |
| .829 | .747 |
| 1.064 | .980 |
| 1.050 | .868 |
| 1.002 | .891 |
| 1.117 | 1.007 |
| .980 | .874 |
| 1.037 | 1.169 |

**두 변수의 평균 비교**




# SPSS를 이용한 대응표본 T 검정

# SPSS 분석 결과

**대응표본 통계량**

| | | 평균 | N | 표준편차 | 평균의 표준오차 |
|---|---|---|---|---|---|
| 대응 1 | L_BMD | 1.08212 | 2461 | .175140 | .003530 |
| | F_BMD | .94447 | 2461 | .137187 | .002765 |

**대응표본 상관계수**

| | | N | 상관계수 | 유의확률 |
|---|---|---|---|---|
| 대응 1 | L_BMD & F_BMD | 2461 | .728 | .000 |

**대응표본 검정**

| | | 대응차 | | | | | t | 자유도 | 유의확률 (양쪽) |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | 차이의 95% 신뢰구간 | | | | |
| | | 평균 | 표준편차 | 평균의 표준오차 | 하한 | 상한 | | | |
| 대응 1 | L_BMD - F_BMD | .137652 | .120432 | .002428 | .132892 | .142413 | 56.702 | 2460 | .000 |

# 보고서의 작성

**Table. Mean Comparison between L-BMD and F-BMD**

| Variables | L-spine BMD | Femur BMD | *t*-value | *p*-value |
|---|---|---|---|---|
| | *n* = 2,461 | *n* = 2,461 | | |
| Total | 1.082 ± 0.175 | 0.944 ± 0.137 | 56.702 | 0.000 |

**요추골밀도와 대퇴골밀도 평균의 차이가 있는지 알아보기 위하여 대응표본 T 검정을 수행하였다. 요추골밀도 평균은 1.082이고, 대퇴골밀도 평균은 0.944인 것으로 나타났다. 그리고 이러한 차이는 유의수준 0.001에서 통계적으로 유의한 것이었다.**

**대응표본 통계량**

| | | 평균 | N | 표준편차 | 평균의 표준오차 |
|---|---|---|---|---|---|
| 대응 1 | L_BMD | 1.08212 | 2461 | .175140 | .003530 |
| | F_BMD | .94447 | 2461 | .137187 | .002765 |

**대응표본 상관계수**

| | | N | 상관계수 | 유의확률 |
|---|---|---|---|---|
| 대응 1 | L_BMD & F_BMD | 2461 | .728 | .000 |

**대응표본 검정**

| | | 대응차 | | | | | t | 자유도 | 유의확률 (양쪽) |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | 차이의 95% 신뢰구간 | | | | |
| | | 평균 | 표준편차 | 평균의 표준오차 | 하한 | 상한 | | | |
| 대응 1 | L_BMD - F_BMD | .137652 | .120432 | .002428 | .132892 | .142413 | 56.702 | 2460 | .000 |

# Act 3

# 독립표본 vs. 대응표본

**독립표본 T 검정(independent sample t-test)**

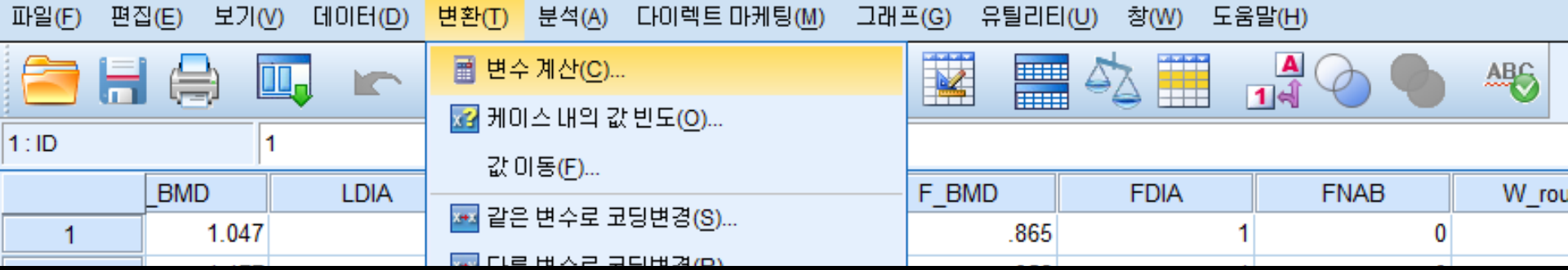
- 독립된 두 집단이 특정 변수의 평균에 있어서 차이가 있는지를 검정하기 위한 것
- 대응표본 T 검정은 하나의 모집단 내에서 처리의 효과를 분석하기 위한 것인 반면, 독립 표본 T 검정은 독립적인 두 집단(실험군과 대조군) 사이의 차이를 분석
- 두 모집단은 정규분포를 따른다는 가정은 필요
- 비교하려는 두 집단은 독립적 임의표본이어야 한다.
  즉 각 대상자는 어느 한 집단에만 속하여야지 두 집단에 모두 속해서는 안 됨

**대응표본 T 검정(paired sample t-test)**

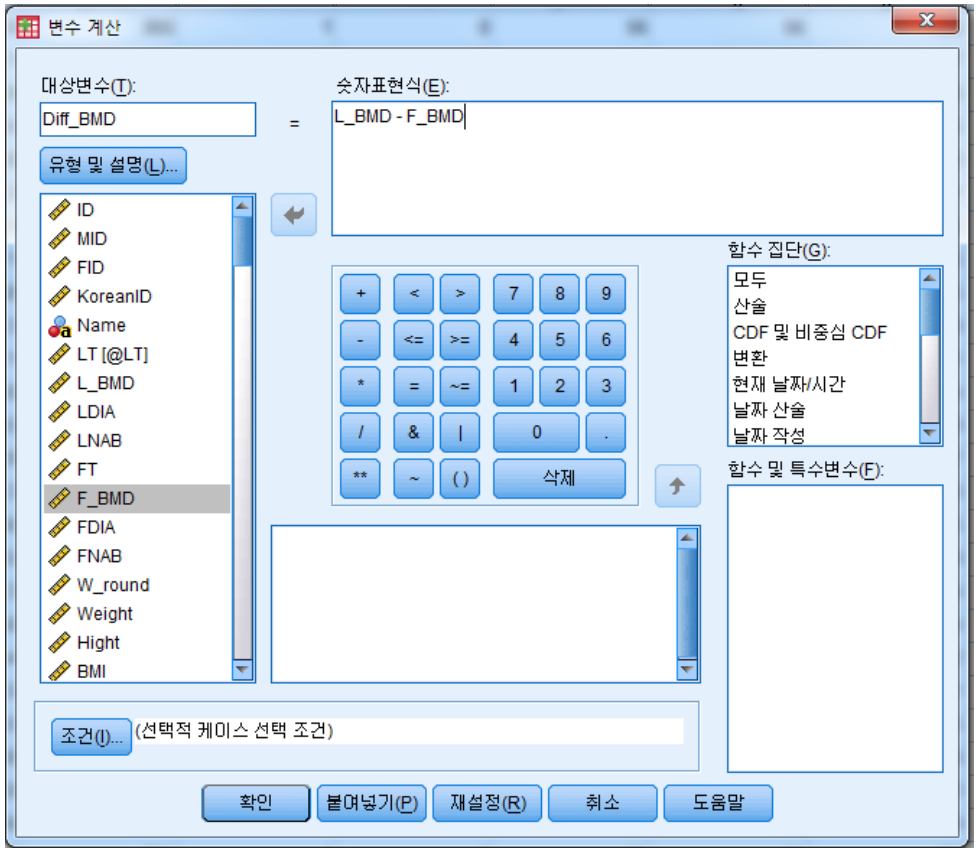
- 한 대상자에게 처치(treatment)를 하기 전과 후의 처치에 효과가 있는지는 검정
- 임상연구에서 치료의 효과가 있는지를 파악하기 위해서 흔히 사용
- 모집단은 정규분포를 따른다는 가정은 필요
- 짝을 이룬 변수는 서로 독립적이지 않으며 하나의 모집단에서 추출된 것

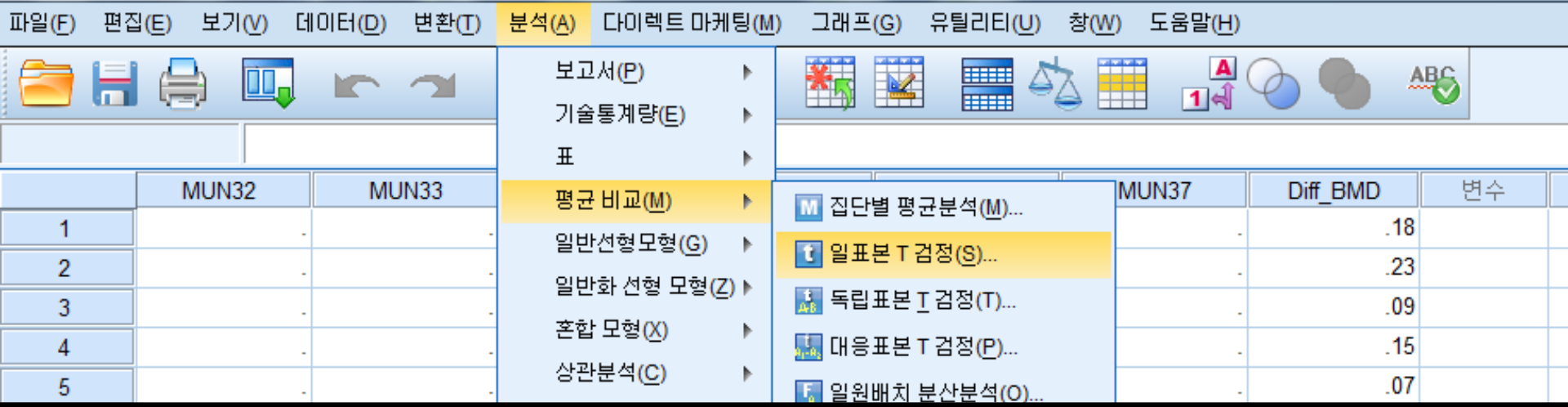
# 주의 사항

짝진 데이터 즉 대응표본을 독립표본으로 간주하여 T-Test를 수행하면 안 되는가? **절대 안 된다.** 이는 대응표본 T 검정의 경우 독립표본 T 검정보다 일반적으로 검정력이 높기 때문이다. 표본이 독립적으로 다루어질 때 무시되었던 정보를 대응표본 T 검정에서는 고려한다는데 그 이유가 있다. 예를 들면, 대응표본 T 검정에서는 각 대상자의 치료 전 혈압이 치료 후에 어떻게 변화하는가를 하나 하나 고려하게 된다. 이처럼 두 표본이 서로 종속적이라는 사실에도 불구하고, 독립표본 T 검정을 수행할 경우 불필요하게 검정력의 손실을 발생시키게 된다. 따라서 두 표본의 데이터가 짝진 경우라면 서로 독립적인 것으로 다루어서는 안 된다!

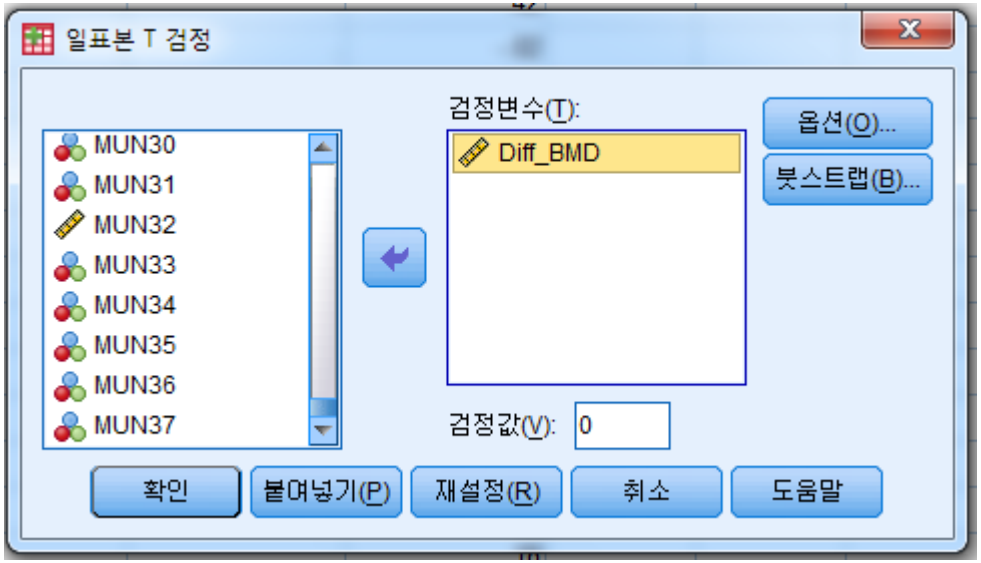



# SPSS를 이용한 변수의 생성






# SPSS를 이용한 일표본 T 검정



# SPSS 분석 결과

**일표본 통계량**

| | N | 평균 | 표준편차 | 평균의 표준오차 |
|---|---|---|---|---|
| Diff_BMD | 2461 | .1377 | .12043 | .00243 |

**일표본 검정**

| | 검정값 = 0 | | | | | |
|---|---|---|---|---|---|---|
| | | | | | 차이의 95% 신뢰구간 | |
| | t | 자유도 | 유의확률 (양쪽) | 평균차 | 하한 | 상한 |
| Diff_BMD | 56.702 | 2460 | .000 | .13765 | .1329 | .1424 |

**대응표본 통계량**

| | | 평균 | N | 표준편차 | 평균의 표준오차 |
|---|---|---|---|---|---|
| 대응 1 | L_BMD | 1.08212 | 2461 | .175140 | .003530 |
| | F_BMD | .94447 | 2461 | .137187 | .002765 |

**대응표본 상관계수**

| | | N | 상관계수 | 유의확률 |
|---|---|---|---|---|
| 대응 1 | L_BMD & F_BMD | 2461 | .728 | .000 |

**대응표본 검정**

| | | 대응차 | | | | | t | 자유도 | 유의확률 (양쪽) |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | 차이의 95% 신뢰구간 | | | | |
| | | 평균 | 표준편차 | 평균의 표준오차 | 하한 | 상한 | | | |
| 대응 1 | L_BMD - F_BMD | .137652 | .120432 | .002428 | .132892 | .142413 | 56.702 | 2460 | .000 |

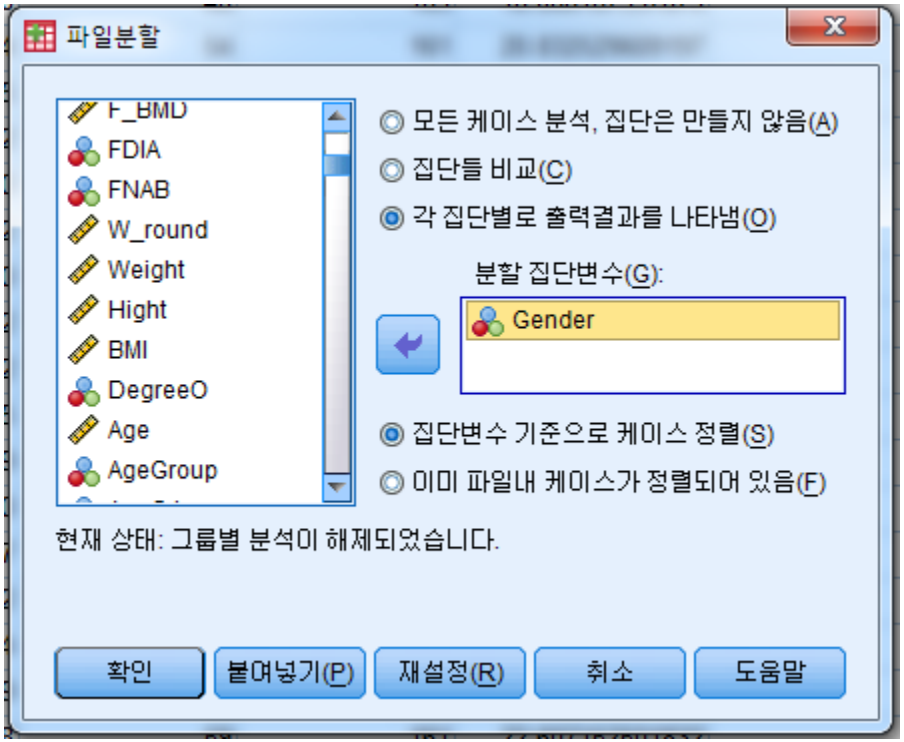
# 집단별 분석

파일(F) 편집(E) 보기(V) 데이터(D) 변환(T) 분석(A) 다이렉트 마케팅(M) 그래프(G) 유틸리티(U) 창(W) 도움말(H)

- 변수 특성 정의(V)...
- 데이터 특성 복사(C)...
- 새 사용자 정의 속성(B)...
- 날짜 정의(A)...
- 다중응답 변수군 정의(M)...
- 확인(L)
- 중복 케이스 식별(U)...
- 특수 케이스 식별(I)...
- 케이스 정렬(O)...
- 변수 정렬(B)...
- 데이터 전치(N)...
- 구조변환(R)...
- 파일 합치기(G)
- 데이터 통합(A)...
- 직교계획(H)
- 데이터 파일 복사(D)
- 파일분할(F)...
- 케이스 선택(S)...
- 가중 케이스(W)...

1 : ID 1

| | FT | FNAB | W_round | Weight | Hight | BM |
|---|---|---|---|---|---|---|
| 1 | -.6 | 0 | 56 | 56 | 156 | 23.0111 |
| 2 | .1 | 0 | 54 | 54 | 157 | 21.9075 |
| 3 | -.9 | 0 | 78 | 78 | 176 | 25.1807 |
| 4 | -1.0 | 0 | 50 | 50 | 165 | 18.3654 |
| 5 | -.8 | 0 | 48 | 48 | 163 | 18.0661 |
| 6 | -.2 | 0 | 54 | 54 | 161 | 20.8325 |
| 7 | -.8 | 0 | 55 | 55 | 164 | 20.4491 |
| 8 | .1 | 0 | 62 | 62 | 155 | 25.8064 |
| 9 | -.8 | 0 | 60 | 60 | 165 | 22.0385 |
| 10 | -.3 | 0 | 52 | 52 | 160 | 20.3125 |
| 11 | .1 | 0 | 60 | 60 | 161 | 23.1472 |
| 12 | -1.0 | 0 | 52 | 52 | 154 | 21.9261 |
| 13 | -1.5 | 1 | 62 | 62 | 159 | 24.5243 |
| 14 | .9 | 0 | 55 | 55 | 157 | 22.3132 |
| 15 | -.3 | 0 | 69 | 69 | 170 | 23.8754 |
| 16 | 1.6 | 0 | 60 | 60 | 155 | 24.9739 |
| 17 | -1.5 | 1 | 47 | 47 | 159 | 18.5910 |
| 18 | .3 | 0 | 52 | 52 | 164 | 19.3337 |
| 19 | -.6 | 0 | 54 | 54 | 154 | 22.7694 |


파일분할
F_BMD
FDIA
FNAB
W_round
Weight
Hight
BMI
DegreeO
Age
AgeGroup
모든 케이스 분석, 집단은 만들지 않음(A)
집단들 비교(C)
각 집단별로 출력결과를 나타냄(O)
분할 집단변수(G):
Gender
집단변수 기준으로 케이스 정렬(S)
이미 파일내 케이스가 정렬되어 있음(F)
현재 상태: 그룹별 분석이 해제되었습니다.
확인
붙여넣기(P)
재설정(R)
취소
도움말