

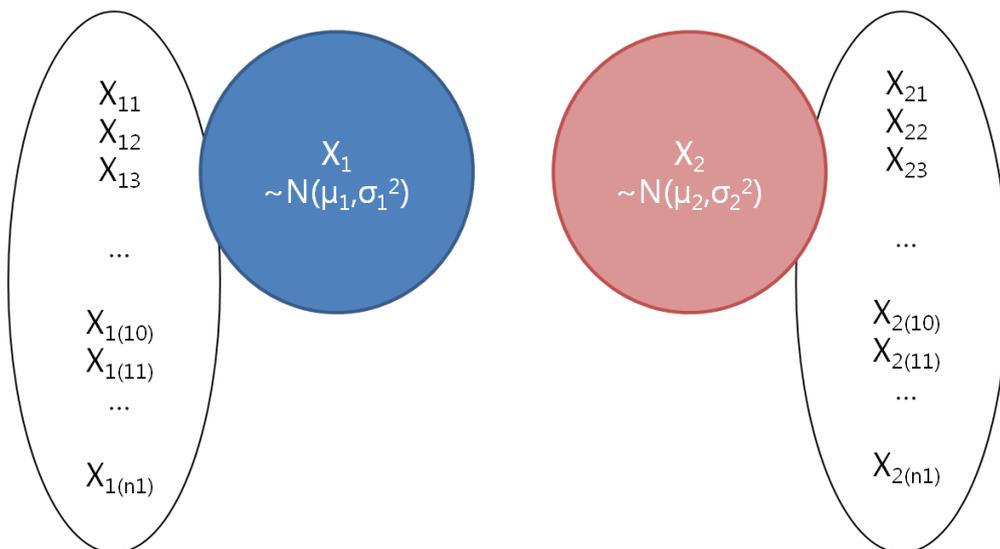
## 독립이표본 t-검정(independent two-samples t-test)

### □ 적용방법

- 1) 성별과 같은 두 독립된 집단의 차이검정은 독립 이표본 t-검증(independent two-samples t-test)를 통해 검증한다.
- 2) 예) 성별에 따른 비만도의 차이, 사망여부에 따라 혈액정보의 차이

### □ 적용 프로세스

- 1) 먼저, 두 집단의 분산의 동일성을 검증하고 난 다음, 두 집단의 평균의 차이를 검증하게 됨.
- 2) 두 집단의 분산이 같은지 **Levene** 의 등분산 검정(Levene's test for equality of variance)을 실행
- 3) 등분산이 검증되면 두 집단의 분산은 동일하다는 뜻이므로 합동표본분산(pooled sample variance)를 차이검정에 적용
- 4) 등분산이 검증되지 않으면 합동표본분산이 아닌 각 집단의 분산을 활용하여 차이검정 실행





□ 독립 이표본 t-검정에서의 독립변수와 종속변수

- 1) T-검정에서의 독립변수는 독립된 범주가 2 개로 구성된 명목형 변수이며, 종속변수는 5 점 척도 또는 연속형 변수로 구성
- 2) 명목형 변수가 연속형 변수에 영향을 미치는지를 검증하게 되며, 명목형 변수는 범주가 2 개인 변수여야 함

□ 검정통계량(등분산 가정)

먼저 두 집단의 분산이 같은지 Levene 의 등분산 검증(Levene's test for equality of variance)을 실행한 다음, 등분산이 검증되면 두 집단의 분산은 동일하다는 뜻이므로 **합동표본분산(pooled sample variance)**를 차이검증에 적용하게 된다. 따라서, 아래의 수식처럼 차이검증을 진행하게 된다.

- 1) 아래의 검정통계량에서와 같이, 두 집단의 분산이 동일하다고 가정될 경우 분석의 효율성을 위해 하나의 분산을 적용하게 되며 그 분산을 "**합동표본분산**"이라고 함
- 2) 등분산이 가정된 경우나 그렇지 않은 경우에 상관없이 t-검정통계량의 분자는 두 집단의 평균차이가 분자로 배치됨

$$S_p^2 = \frac{\sum_{i=1}^{n_1} (X_i - \bar{X})^2 + \sum_{i=1}^{n_2} (Y_i - \bar{Y})^2}{(n_1 - 1) + (n_2 - 1)}$$

( $S_p^2$  는 합동표본분산,  $\bar{X}$ 는 특정집단의 평균,  $\bar{Y}$ 는 다른 집단의 평균,  $n_1$ 는 특정집단의 표본 수,  $n_2$ 는 다른 집단의 표본 수)

$$\frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2 = 0)}{S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t(n_1 + n_2 - 2)$$



□ **검정통계량(이분산 가정)**

두 집단의 등분산 검증에서 두 집단의 분산이 동일하지 않다고 검증되면 합동표본분산이 아닌 각 집단의 분산을 활용하여 차이검증을 하게 된다. 차이를 검증하고자 하는 두 집단의 분산이 동일분산이 아닐 경우는 다음의 검정 통계량(test statistics)을 적용하게 된다. 두 표본분산이 동일하다는 가정을 할 수 없는 경우, 새터스웨이트(Satterthwaite)의 근사적 방법을 통해 차이검증을 실시하게 된다.

- 1) 두 집단의 분산이 동일한지에 대한 검증에서 이분산 즉 두 집단의 분산이 다르다고 검증될 경우, 각 집단의 분산을 검정통계량에 적용하게 됨
- 2) 등분산이 가정된 경우나 그렇지 않은 경우에 상관없이 t-검정통계량의 분자는 두 집단의 평균차이가 분자로 배치됨

$$\frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2 = 0)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \sim t(v)$$

$$v = \frac{(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2})^2}{\frac{S_1^4}{n_1^2(n_1 - 1)} + \frac{S_2^4}{n_2^2(n_2 - 1)}} - 2$$

( $S_1^2$  는 특정집단의 분산,  $S_2^2$  는 다른 집단의 분산,  $\bar{X}$  는 특정집단의 평균,  $\bar{Y}$  는 다른 집단의 평균,

$n_1$  는 특정집단의 표본 수,  $n_2$  는 다른 집단의 표본 수)



## □ t-검정통계량과 p-value 의 관계

### 1) t-값 $\uparrow$ $\rightarrow$ 두 집단의 평균차이 $\uparrow$ $\rightarrow$ p-value $\downarrow$

- i. 검정통계량의 값이 높아지는 것은  
검정통계량의 분자인 두 집단의 평균차이가 상대적으로 크다는 것을 의미함
- ii. 또는 검정통계량의 분모가 상대적으로 굉장히 작음을 의미함

### 2) t-값 $\downarrow$ $\rightarrow$ 두 집단의 평균차이 $\downarrow$ $\rightarrow$ p-value $\uparrow$

- i. 검정통계량의 값이 낮아지는 것은  
검정통계량의 분자인 두 집단의 평균차이가 상대적으로 작다는 것을 의미함
- ii. 또는 검정통계량의 분모가 상대적으로 높음을 의미함

### Main Tip> t-값의 높고 낮음

독립 이표본 t-검정에서는 t-검정통계량을 기초로 하여 검정하게 되는데, t-값이 높을 경우 두 집단의 평균차이가 큰 것을 의미하며 두 집단에 따른 연속형 변수의 평균값이 차이가 큰 것을 의미한다.

**결과적으로 독립변수는 2개의 범주로 구성된 명목형 변수이며, 종속변수는 연속형 변수 일 때 독립된 두 집단의 평균차이가 클수록 통계적으로 유의할 확률이 높아진다.**



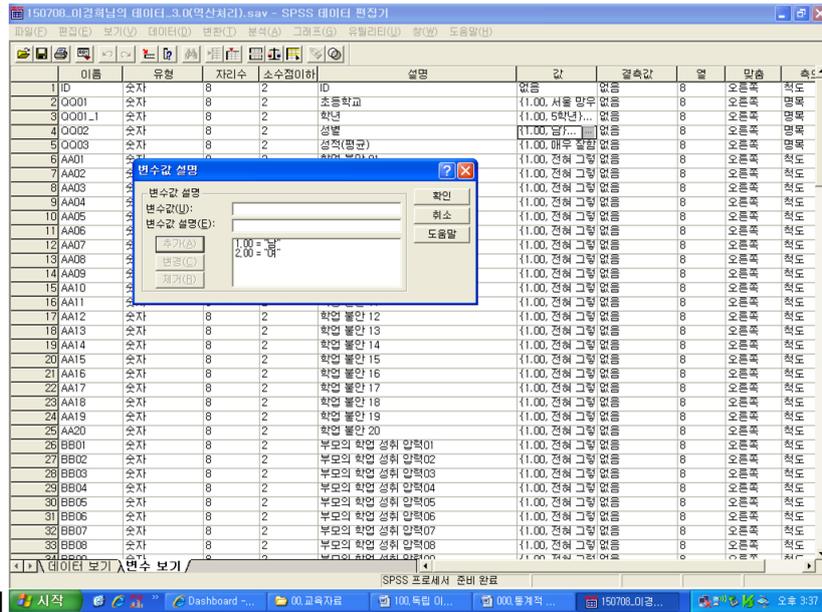
## C Table 생성에 대한 실습

### □ C Table 의 유용성

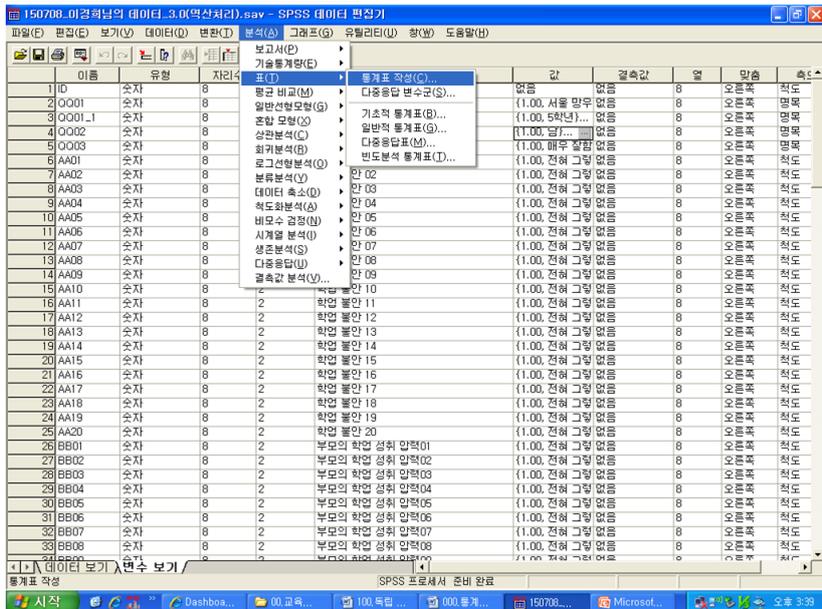
- 1) Column 기준의 테이블을 생성하는 것으로  
카이제곱 독립성 검정 및 집단간 차이검정의 결과를 효과적으로 표현할 수 있음.
- 2) 논문상에서 두 집단간 차이검정에서 C Table 양식으로 정리하는 것이 일반적임.

□ C Table 생성방법

1) 먼저 C Table 을 생성할 데이터를 생성

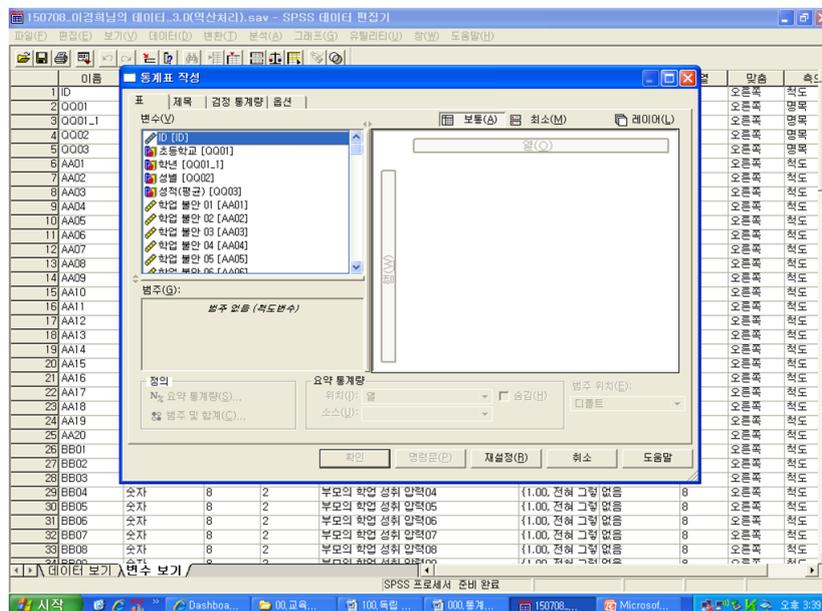


2) 분석 ▶ 표 ▶ 통계표 작성을 선택한다.



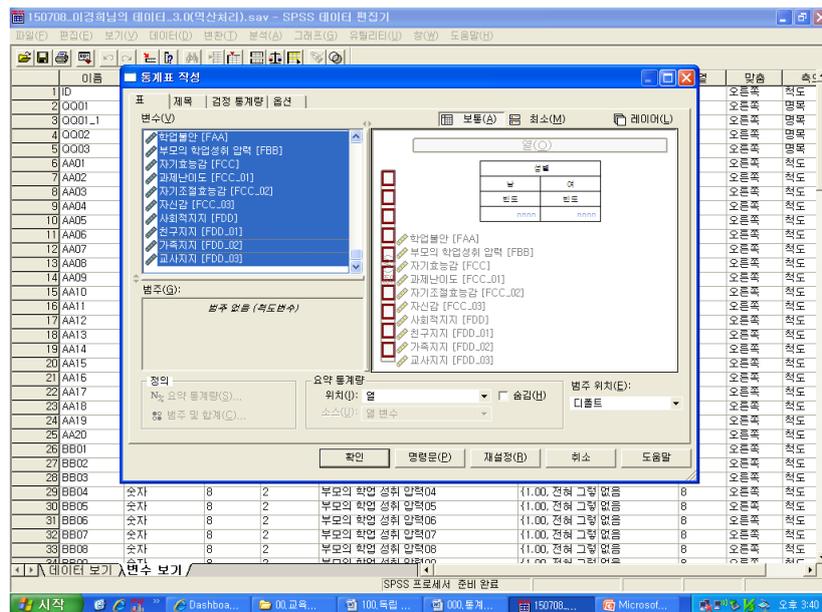
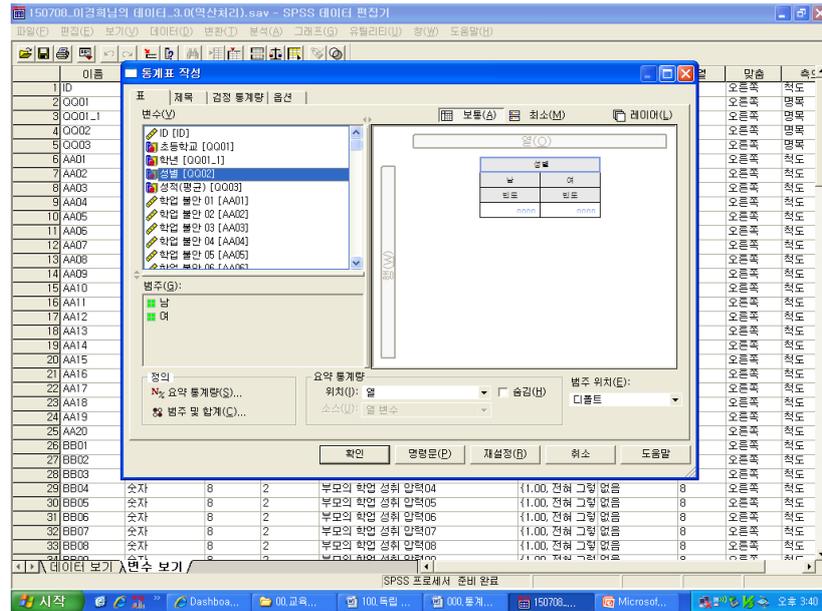
3) 통계표 작성이라는 대화상자 활성화

- i. 왼쪽에서 선택 가능한 변수들이 나열되어 있으며, 오른쪽에는 행과 열의 배치를 확인할 수 있다.
- ii. C Table 생성을 위한 선택조건
  - ⇒ 명목형 변수는 명목형으로 설정해야 하며, 연속형 변수는 연속형으로 설정해야 함



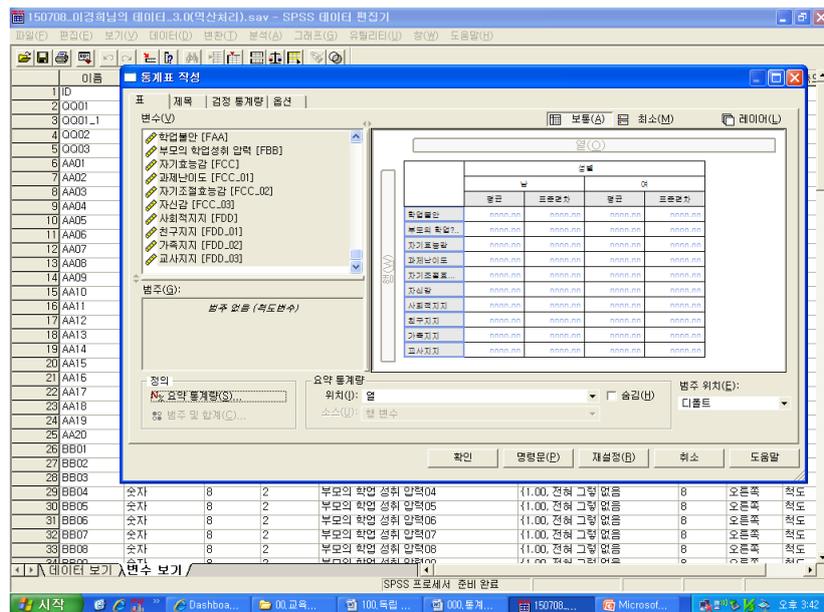
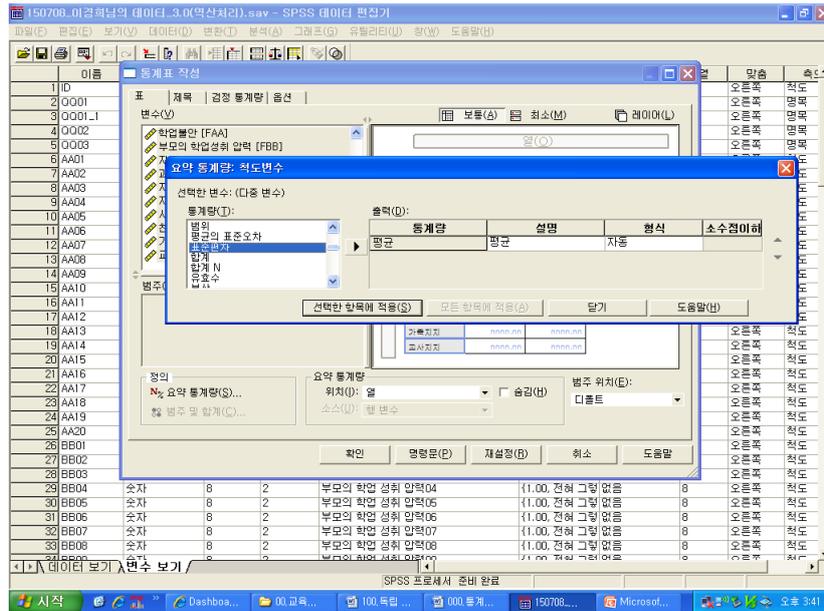
4) 그룹 변수를 선정

- i. 일반적으로 그룹 변수를 열 위치에 옮겨 놓는다.
- ii. 연속형 변수는 아래와 같이 행 위치에 배치한다.



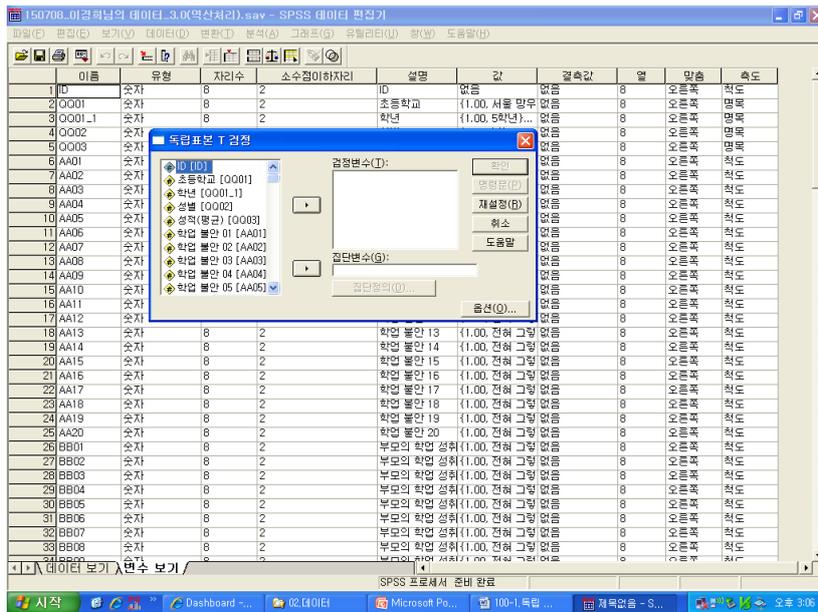
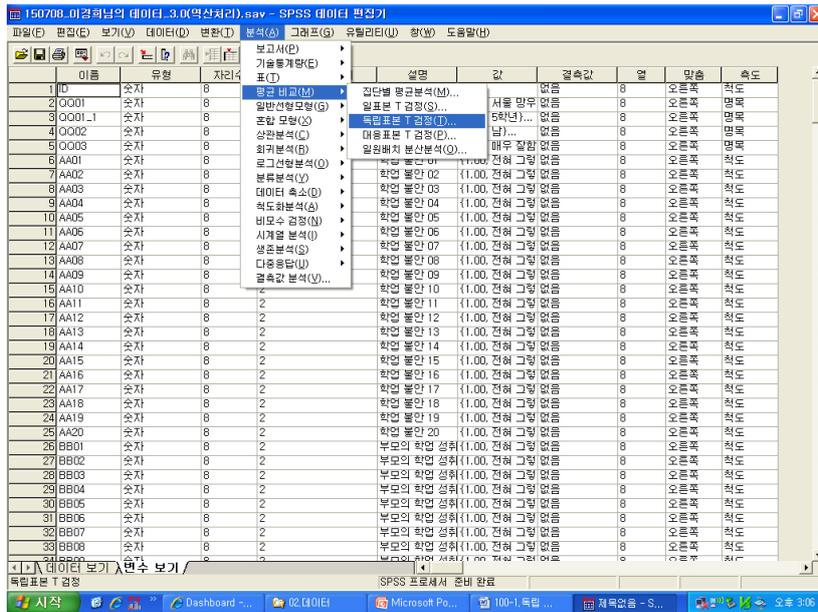
5) 표준편차 지정 및 Output Image 최종결정

- i. 왼쪽 하단의 “요약 통계량”을 선택하면, 평균이 디폴트로 설정되어 있음
- ii. 표준편차를 선택한 다음, “선택한 항목에 적용”을 클릭하여 종료함



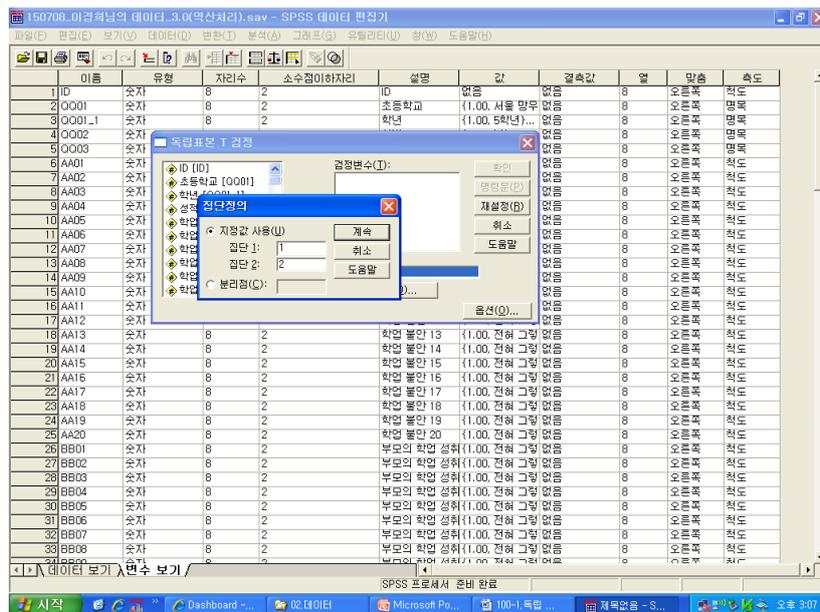
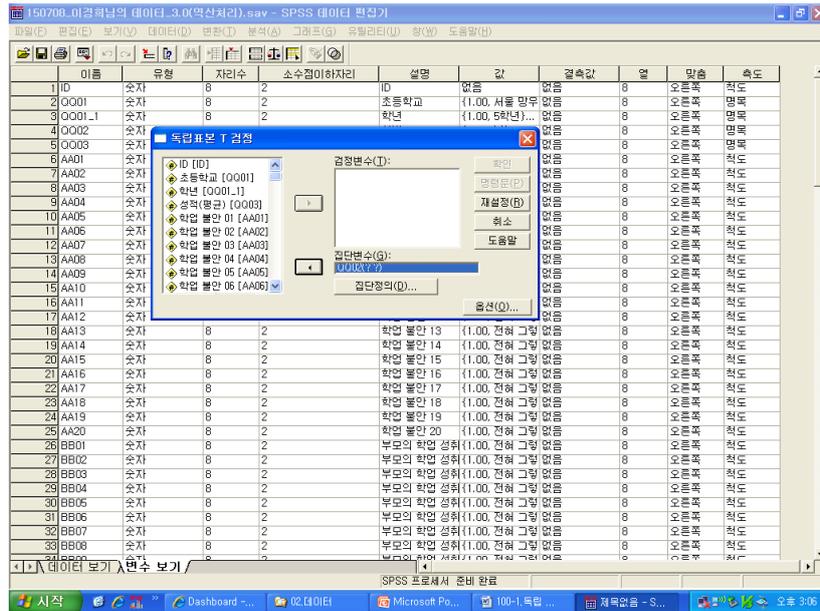
## 독립이표본 t-검정에 대한 실습

### 1) 분석 ▶ 평균비교 ▶ 독립표본 T 검정을 선택한다



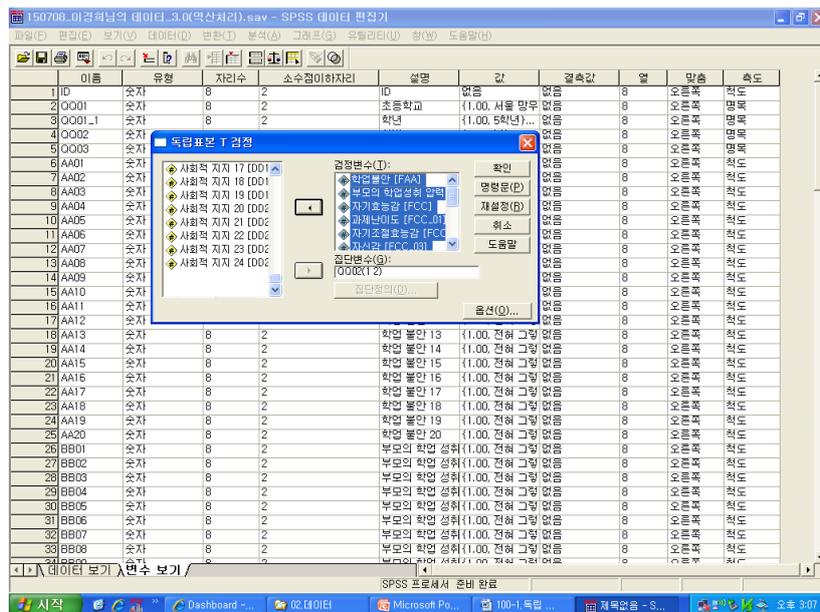
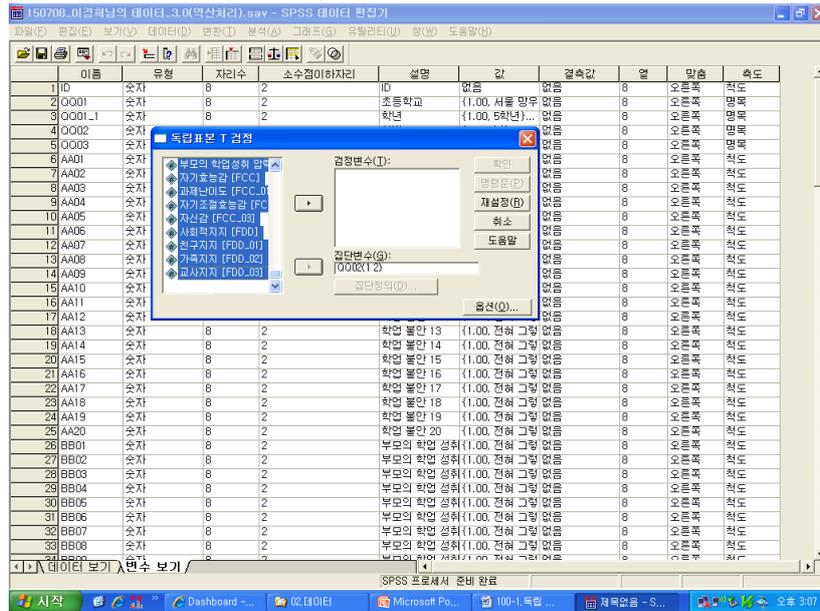
2) 집단변수 선정

- i. 집단변수에 독립된 두 개의 범주로 구성된 명목형 변수를 지정함
- ii. 명목형 변수의 범주 지정값을 지정함



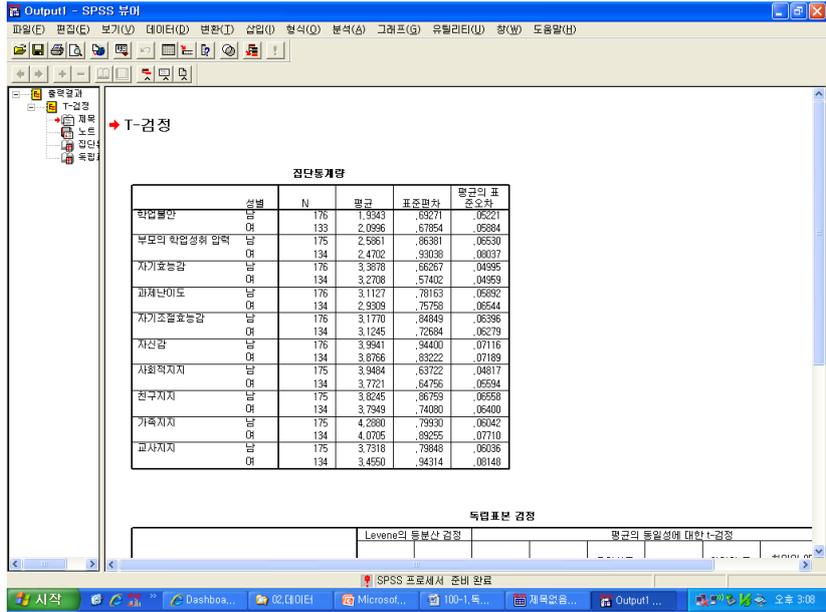
3) 검정변수 지정

- i. 검정변수에는 연속형 변수를 지정하게 된다.
- ii. 다음과 같이, 연속형 변수를 검정변수 란으로 이동한다.



4) 분석결과 도출 및 해석

- i. T-검정에서는 집단별 연속형 변수에 대한 기술통계가 도출되며,
- ii. 독립표본 검정에서는 두 집단의 등분산 검정과 t-검정결과를 동시에 제공

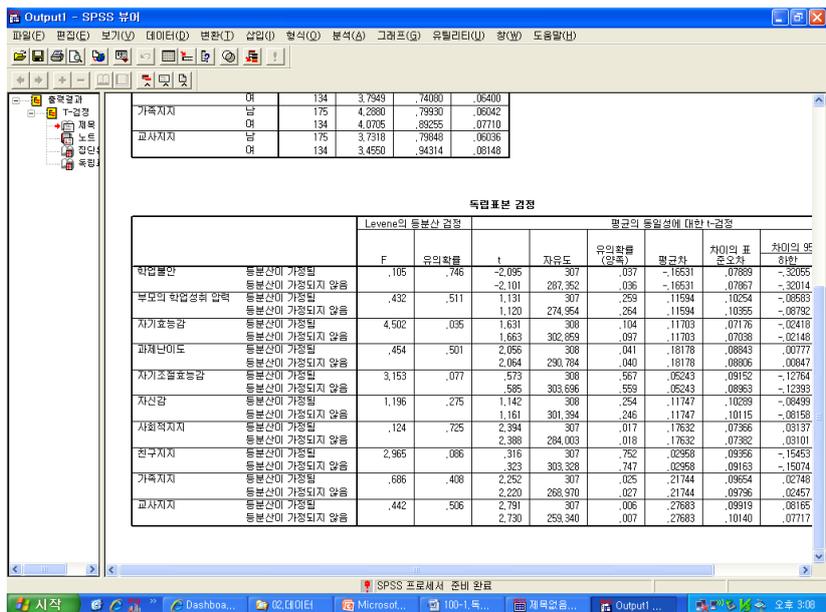


**집단통계량**

변수	성별	N	평균	표준편차	평균의 표준오차
학업불만	남	176	1.9343	.69271	.05221
	여	133	2.0996	.67854	.05884
부모의 학업성취 압력	남	175	2.5661	.86381	.06630
	여	134	2.4702	.93036	.08037
자기효능감	남	176	3.3878	.96267	.04995
	여	134	3.2708	.57402	.04959
과제난이도	남	176	3.1127	.78163	.05892
	여	134	2.9308	.75758	.05544
자기조절효능감	남	176	3.1770	.84949	.06356
	여	134	3.1245	.72684	.06279
자신감	남	176	3.9941	.94400	.07116
	여	134	3.8766	.83222	.07189
사회적지지	남	175	3.9484	.83722	.04817
	여	134	3.7721	.84756	.05594
친구지지	남	175	3.8246	.86759	.06558
	여	134	3.7949	.74080	.06400
가족지지	남	175	4.2880	.79930	.06042
	여	134	4.0705	.89295	.07710
교사지지	남	175	3.7318	.79948	.06036
	여	134	3.4550	.94314	.08148

**독립표본 검정**

Levene의 등분산 검정	평균의 동일성에 대한 T-검정



**독립표본 검정**

변수	Levene의 등분산 검정	평균의 동일성에 대한 T-검정							
		F	유의확률	t	자유도	유의확률 (양측)	평균차	차이의 표준오차	차이의 95% 하한
학업불만	등분산이 가정됨 등분산이 가정되지 않음	.105	.746	-2.095	307	.037	-.16531	.07889	-.30955
부모의 학업성취 압력	등분산이 가정됨 등분산이 가정되지 않음	.432	.511	1.131	287.352	.036	-.16531	.07887	-.30914
자기효능감	등분산이 가정됨 등분산이 가정되지 않음	4.502	.035	1.631	308	.104	.11703	.07176	-.02418
과제난이도	등분산이 가정됨 등분산이 가정되지 않음	.454	.501	1.663	302.859	.097	.11703	.07038	-.02148
자기조절효능감	등분산이 가정됨 등분산이 가정되지 않음	3.153	.077	2.064	290.784	.041	.18178	.08843	.03777
자신감	등분산이 가정됨 등분산이 가정되지 않음	1.196	.275	1.142	308	.254	.11747	.10289	-.08499
사회적지지	등분산이 가정됨 등분산이 가정되지 않음	.124	.725	2.394	307	.017	.17632	.07366	.03137
친구지지	등분산이 가정됨 등분산이 가정되지 않음	2.965	.086	3.16	307	.006	.02958	.08956	-.15453
가족지지	등분산이 가정됨 등분산이 가정되지 않음	.686	.408	2.252	307	.025	.21744	.08654	.02746
교사지지	등분산이 가정됨 등분산이 가정되지 않음	.442	.506	2.791	307	.006	.27683	.09919	.08165