**데이터 분석 과제**

전성민, 2022-12-20

## 1. 목표

* 두집단의 중도탈락의도, 학업스트레스, 사회적지지, 자기효능감, 대학만족에 대한 t-검정 실시한 후 분석과 결과에 대한 의미를 작성

\*\* 남학생 vs 여학생, 저학년(1,2학년) vs 고학년 (3학년 이상)

* 모든 변수 (중도탈락의도, 학업스트레스, 사회적지지, 자기효능감, 대학만족) 간 피어슨 상관분석을 실시한 후 분석 결과의 의미를 작성

## 2. 가설 수립

### 2.1 평균 검증

* 귀무가설 (H0) : 중도탈락의도, 학업스트레스, 사회적지지도, 자기효능감, 대학만족 은 남학생/여학생, 저학년/고학년 간 (평균이) 동일할 것이다
* 대립가설 (H1) : 중도탈락의도, 학업스트레스, 사회적지지도, 자기효능감, 대학만족 은 남학생/여학생, 저학년/고학년 간 (평균이) 동일하지 않을 것이다

### 2.2 상관 분석

* 귀무가설 (H0) : 중도탈락의도, 학업스트레스, 사회적지지도, 자기효능감, 대학만족 은 상호관계가 독립적일 것이다
* 대립가설 (H1) : 중도탈락의도, 학업스트레스, 사회적지지도, 자기효능감, 대학만족 은 상호관계가 독립적이지 않을 것이다

### 2.3 유의 수준 (p-value) : 0.05

## 3. 데이터 탐색

### 3.1 기술통계량

#### 3.1.1 남학생 데이터 기술통계량

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | 학업스트레스 | 사회적지지 | 자기효능감 | 중도탈락의도 | 대학만족 |
| count | 89 | 89 | 89 | 89 | 89 |
| mean | 3.27191 | 4.896629 | 4.626966 | 1.882022 | 4.047191 |
| std | 1.125024 | 0.761644 | 0.731866 | 0.89434 | 0.60153 |
| min | 1 | 2.6 | 2.8 | 1 | 2.2 |
| 25% | 2.6 | 4.6 | 4 | 1.25 | 3.8 |
| 50% | 3.4 | 5 | 4.8 | 1.75 | 4 |
| 75% | 4 | 5.4 | 5 | 2.25 | 4.4 |
| max | 6 | 6 | 6 | 5.25 | 6 |

#### 3.1.2 여학생 데이터 기술통계량

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | 학업스트레스 | 사회적지지 | 자기효능감 | 중도탈락의도 | 대학만족 |
| count | 140 | 140 | 140 | 140 | 140 |
| mean | 3.554286 | 4.91 | 4.384286 | 1.948214 | 3.717143 |
| std | 1.00798 | 0.812253 | 0.751044 | 0.80102 | 0.653485 |
| min | 1 | 2 | 2.2 | 1 | 2 |
| 25% | 2.8 | 4.4 | 3.95 | 1.25 | 3.35 |
| 50% | 3.4 | 5 | 4.4 | 2 | 3.8 |
| 75% | 4.2 | 5.4 | 5 | 2.5 | 4.2 |
| max | 6 | 6 | 6 | 4.5 | 5.2 |

#### 3.1.3 저학년 (1,2학년) 데이터 기술통계량

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | 학업스트레스 | 사회적지지 | 자기효능감 | 중도탈락의도 | 대학만족 |
| count | 108 | 108 | 108 | 108 | 108 |
| mean | 3.505556 | 4.794444 | 4.412963 | 2.078704 | 3.738889 |
| std | 0.986434 | 0.832723 | 0.724613 | 0.832775 | 0.670588 |
| min | 1 | 2 | 2.6 | 1 | 2.2 |
| 25% | 3 | 4.4 | 3.95 | 1.5 | 3.2 |
| 50% | 3.4 | 5 | 4.5 | 2 | 3.8 |
| 75% | 4 | 5.2 | 5 | 2.5 | 4.2 |
| max | 6 | 6 | 6 | 4.5 | 5.2 |

#### 3.1.4 고학년(3학년 이상) 데이터 기술통계량

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | 학업스트레스 | 사회적지지 | 자기효능감 | 중도탈락의도 | 대학만족 |
| count | 121 | 121 | 121 | 121 | 121 |
| mean | 3.390083 | 5.003306 | 4.53719 | 1.783058 | 3.940496 |
| std | 1.125715 | 0.742062 | 0.772888 | 0.819643 | 0.623776 |
| min | 1 | 3 | 2.2 | 1 | 2 |
| 25% | 2.6 | 4.6 | 4 | 1 | 3.6 |
| 50% | 3.4 | 5 | 4.6 | 1.5 | 4 |
| 75% | 4.2 | 5.6 | 5 | 2 | 4.4 |
| max | 6 | 6 | 6 | 5.25 | 6 |

### 3.3 평균 시각화 (포인트 플롯, point plot)

평균값을 점으로 표현하고 평균에 대한 95% 신뢰구간을 차트로 표시

#### 3.2.1 학업스트레스

|  |  |
| --- | --- |
| C:\Users\전정신\Pictures\다운로드.png | C:\Users\전정신\Pictures\다운로드 (1).png |

#### 3.2.2 사회적지지

|  |  |
| --- | --- |
| C:\Users\전정신\Pictures\다운로드 (2).png | C:\Users\전정신\Pictures\다운로드 (3).png |

#### 3.2.3 자기효능감

|  |  |
| --- | --- |
| C:\Users\전정신\Pictures\다운로드 (4).png | C:\Users\전정신\Pictures\다운로드 (5).png |

#### 3.2.4 중도탈락의도

|  |  |
| --- | --- |
| C:\Users\전정신\Pictures\다운로드 (6).png | C:\Users\전정신\Pictures\다운로드 (7).png |

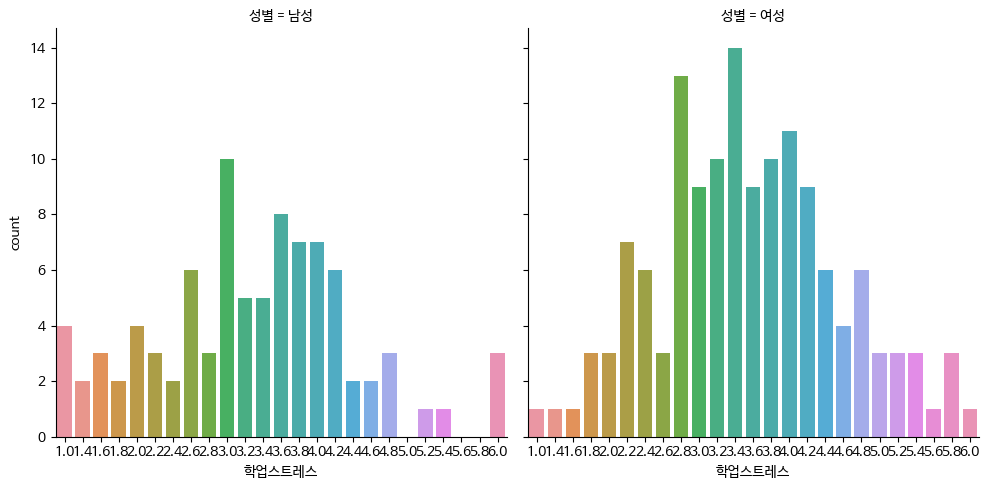
#### 3.2.5 대학만족

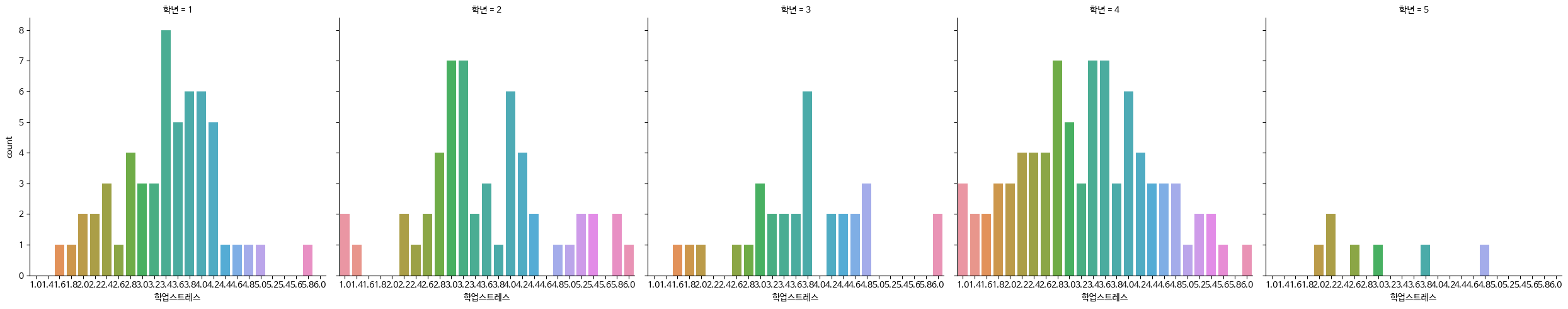
|  |  |
| --- | --- |
| C:\Users\전정신\Pictures\다운로드 (8).png | C:\Users\전정신\Pictures\다운로드 (9).png |

### 3.2 빈도수 시각화

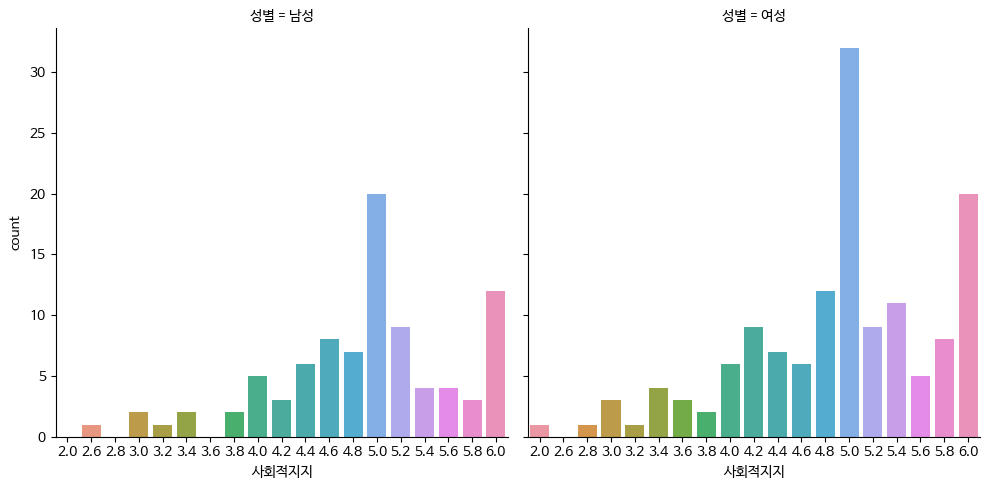
분포 빈도수를 시각적으로 파악

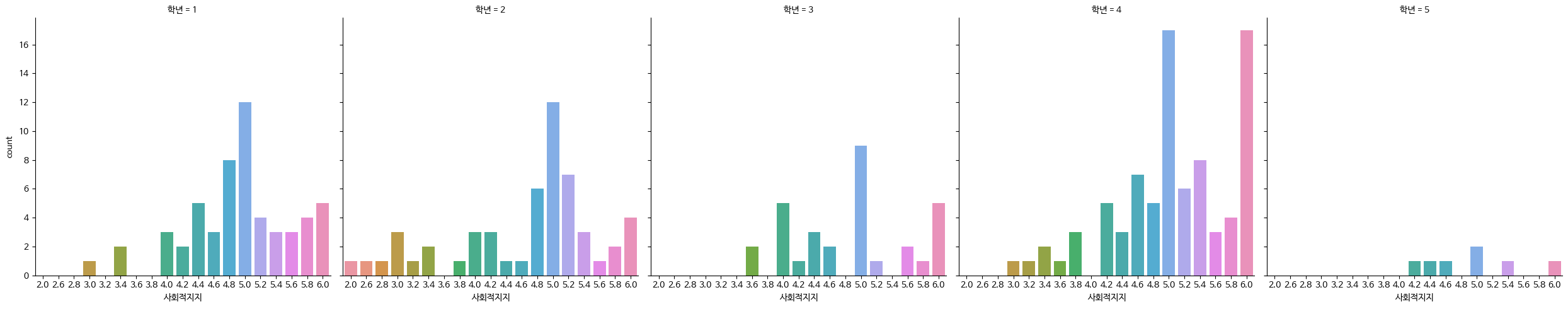
#### 3.2.1 학업스트레스



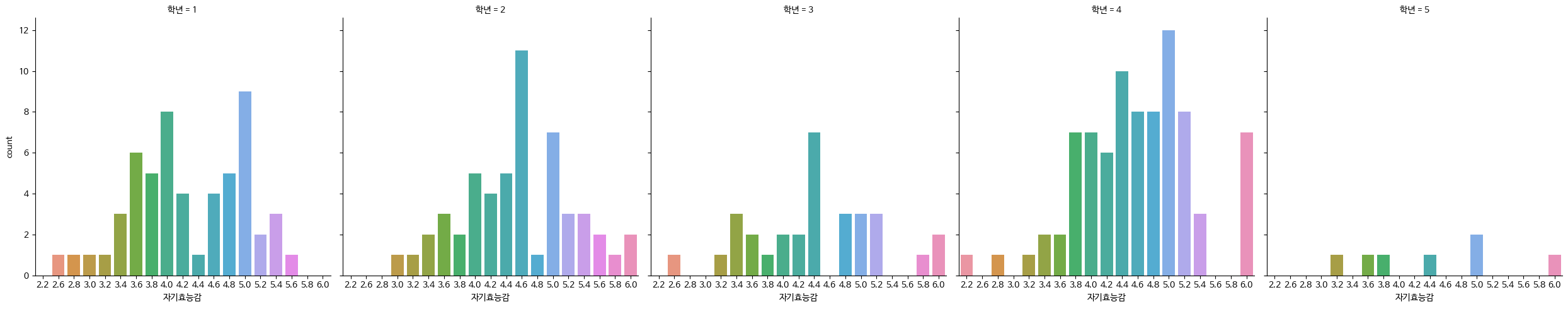
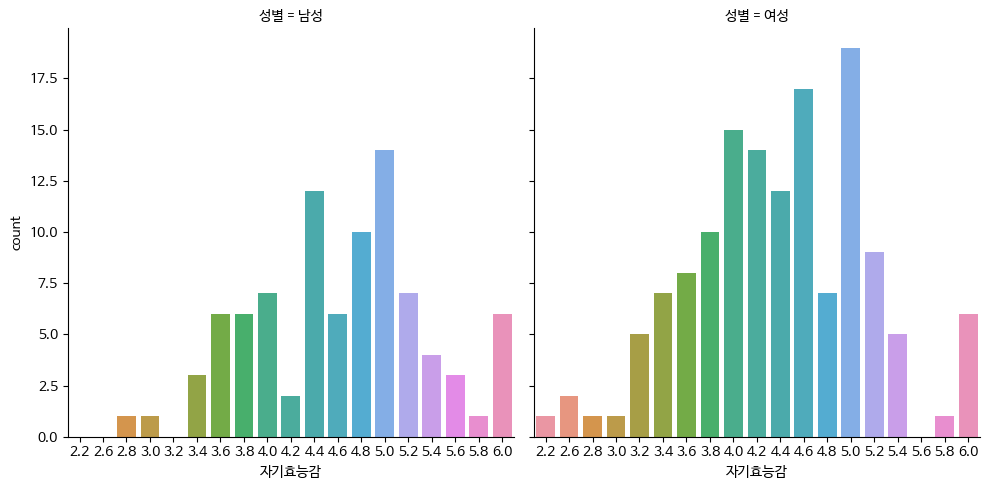


#### 3.2.2 사회적지지

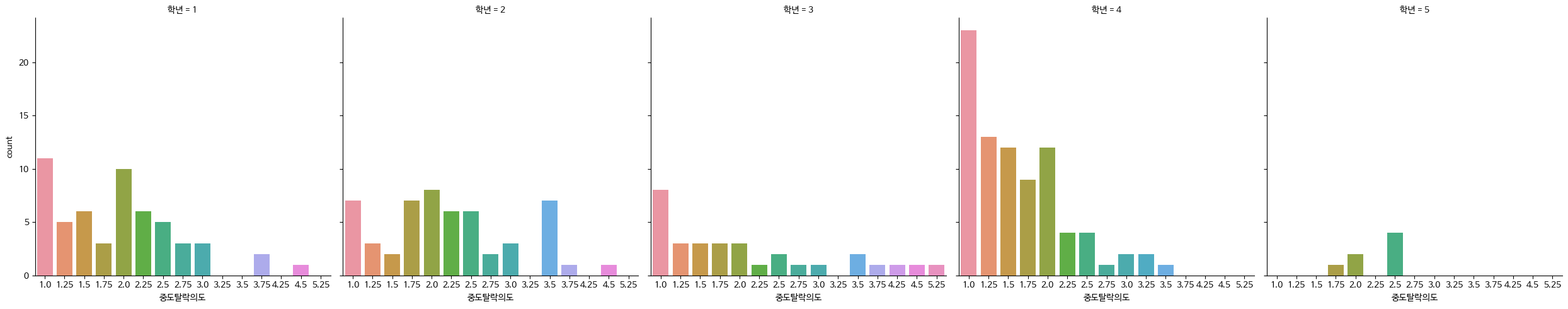
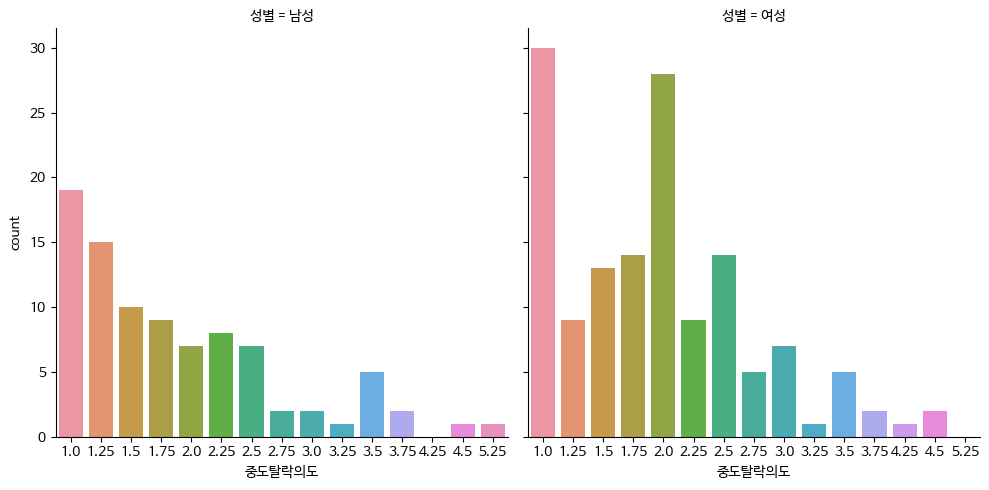




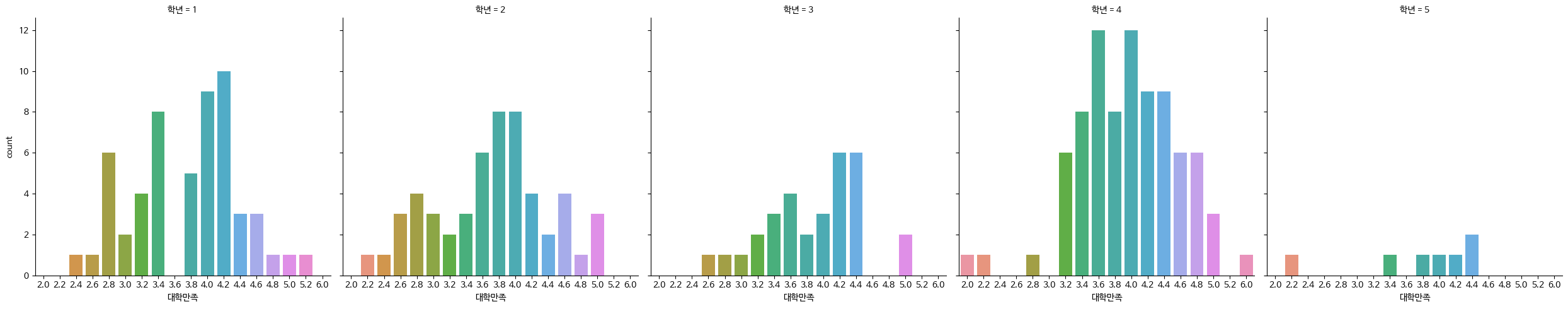
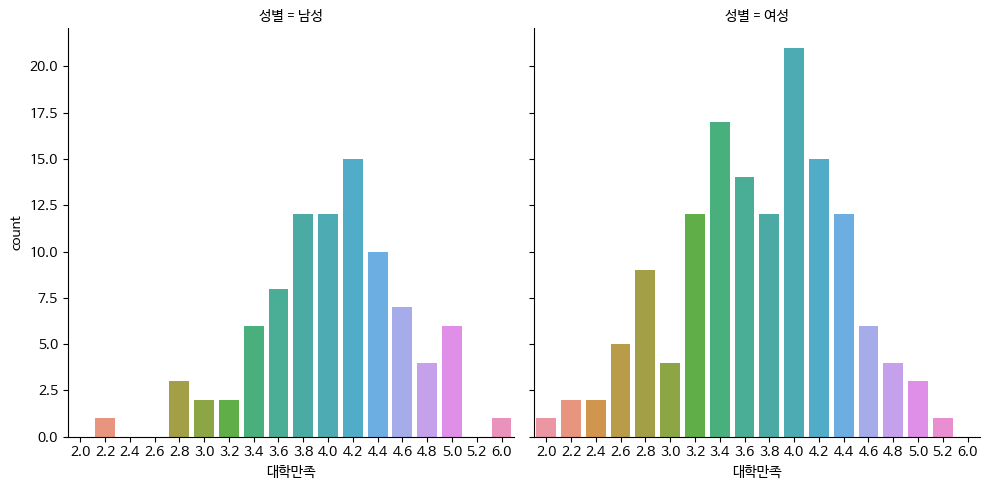
#### 3.2.3 자기효능감



#### 3.2.4 중도탈락의도

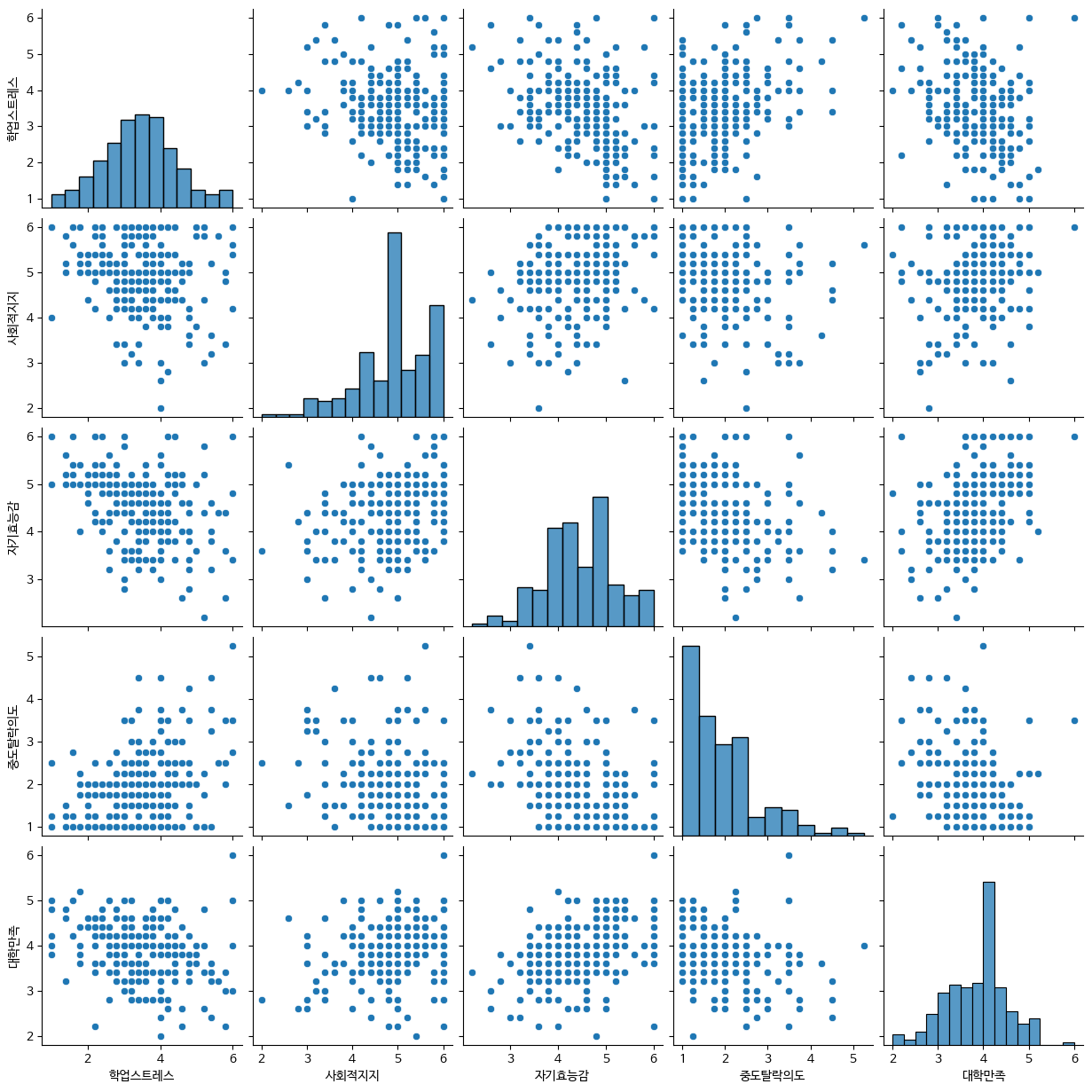


#### 3.2.5 대학만족



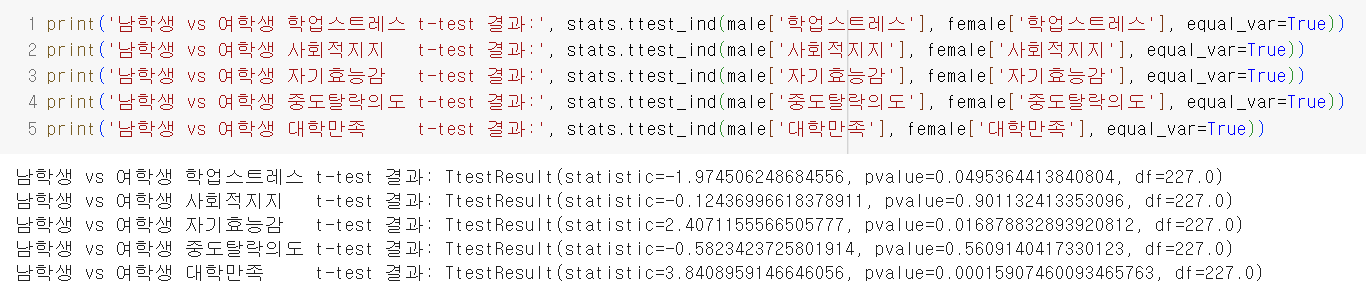
### 3.4 상관관계 시각화 (산점도 행렬, scatterplot matrix)

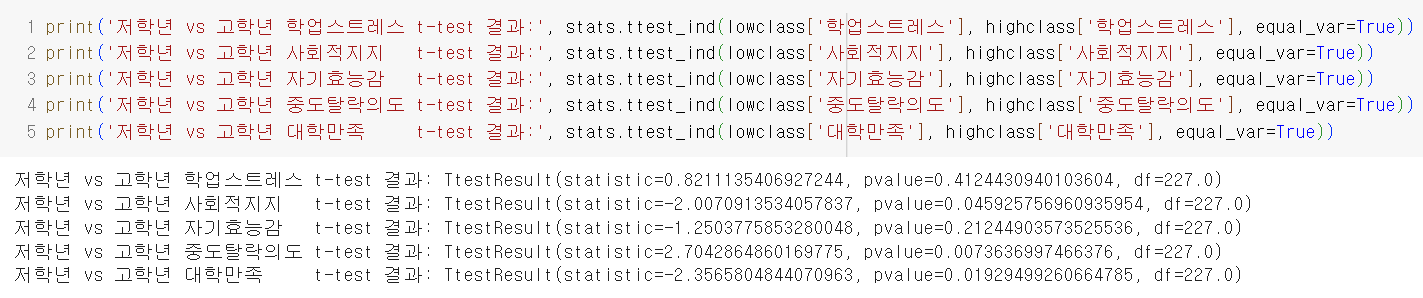
상관계수를 파악하기 전에 산점도를 그려 변수 간에 관련성을 시각적으로 파악



## 4. 데이터 분석결과 및 해석

### 4.1 t-검정 실시 (독립표본 t검정)





[결과 요약]

|  |  |  |
| --- | --- | --- |
| pvalue | 남학생 vs 여학생 | 저학년 vs 고학년 |
| 학업스트레스 | 0.0495364413840804 | 0.4124430940103604 |
| 사회적지지도 | 0.901132413353096 | 0.045925756960935954 |
| 자기효능감 | 0.016878832893920812 | 0.21244903573525536 |
| 중도탈락의도 | 0.5609140417330123 | 0.0073636997466376 |
| 대학만족 | 0.00015907460093465763 | 0.01929499260664785 |

두 집단의 평균 차이는 통계적으로 유의미함 (p<0.05)

[결과 해석]

- 남학생 vs 여학생 표본간

학업스트레스에 유의미한 차이가 조금 있으며, 자기효능감, 대학만족에 느끼는 차이가 많은 것으로 추정됨

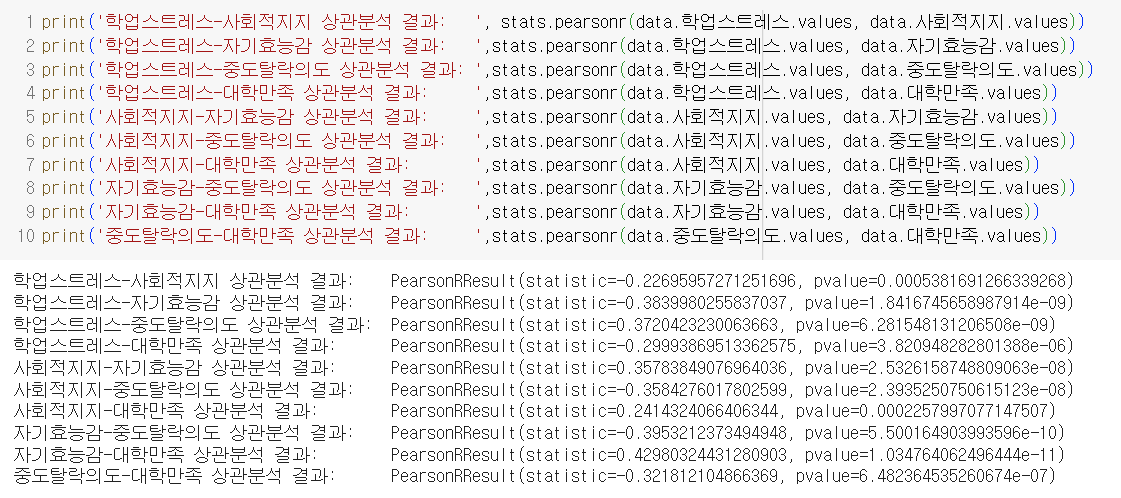
사회적지지도, 중도탈락의도는 유의미한 차이가 없는 것으로 추정

- 저학년 vs 고학년 표본간

사회적지지도에 유의한 차이가 조금 있으며, 중도탈락의도, 대학만족에 느끼는 차이가 많은 것으로 추정됨

학업스트레스, 자기효능감은 유의미한 차이가 없는 것으로 추정

### 4.2 상관분석 실시 (피어슨 상관분석)



[결과 요약]

|  |  |  |
| --- | --- | --- |
|  | statistic | pvalue |
| 학업스트레스-사회적지지 | -0.22695957271251696 | 0.0005381691266339268 |
| 학업스트레스-자기효능감 | -0.3839980255837037 | 1.8416745658987914e-09 |
| 학업스트레스-중도탈락의도 | 0.3720423230063663 | 6.281548131206508e-09 |
| 학업스트레스-대학만족 | -0.29993869513362575 | 3.820948282801388e-06 |
| 사회적지지-자기효능감 | 0.35783849076964036 | 2.5326158748809063e-08 |
| 사회적지지-중도탈락의도 | -0.3584276017802599 | 2.3935250750615123e-08 |
| 사회적지지-대학만족 | 0.2414324066406344 | 0.0002257997077147507 |
| 자기효능감-중도탈락의도 | -0.3953212373494948 | 5.500164903993596e-10 |
| 자기효능감-대학만족 | 0.42980324431280903 | 1.034764062496444e-11 |
| 중도탈락의도-대학만족 | -0.321812104866369 | 6.482364535260674e-07 |

[결과 해석]

상관계수 (correlation coefficient) 두 변수가 함께 변하는 정도를 -1 ~ 1 범위의 수로 나타낸 것으로

상관이 +인 경우 두 변수가 같은 방향으로 변화 (하나가 증가하면 다른 하나도 증가),

상관이 -인 경우 두 변수가 반대 방향으로 변화 (하나가 증가하면 다른 하나는 감소),

상관 = 0이면 두 변수가 독립, 즉, 한 변수의 변화로 다른 변수의 변화를 예측하지 못하며

상관이 클 수록 두 변수는 함께 많이 변화



상관분석 결과 모든 변수의 pvalue가 0.05보다 작아 통계적으로 유의미한 결과이다.

1)학업스트레스-사회적지지, 2)학업스트레스-대학만족 간 약한 음적 상관관계이며,

1)학업스트레스-자기효능감, 2)사회적지지-중도탈락의도, 3)자기효능감-중도탈락의도, 4)중도탈락의도-대학만족 변수간 뚜렷한 음적 상관관계를 확인 할 수 있다. 이는 중도탈락의도 높으면 사회적지지, 자기효능감, 대학만족이 낮은 것을 추정할 수 있다.

1)사회적지지-대학만족 간 약한 양적 상관관계이며,

1)학업스트레스-중도탈락의도, 2)사회적지지-자기효능감, 3)자기효능감-대학만족 변수간 뚜렷한 양적 상관관계를 확인 할 수 있다. 이는 자기효능감이 높을수록 사회적지지와 대학만족이 높은 것을 추정할 수 있다.