



# 경영과학 **PBL** 모듈 1

6조

---

김수연 문지용 박준희 이주연  
이채은 장명은 최윤석

---

# Contents

---

- 1 문제설명 및 목표 설정
- 2 가설 설정
- 3 자료 수집
- 4 상관관계 분석
- 5 선형회귀분석

# Contents

6 한계

7 한계 극복을 위한 방법 제시

8 결론

# 1. 문제 설명 및 목표 설정

## 문제 설명

고등학생을 포함한 대학 진학 예정자가 지원할 대학을 결정할 때 참고할 수 있는 대학 순위 개발

## 목표

대학의 순위를 나열할 수 있는 ‘**객관적 지표**’를 **과학적 방법**을 통해 설정  
(단, 임의적으로 대학의 범위를 **서울소재 4년제 종합대학**으로 한정)

## 2. 가설 설정

### 가설1

대학 진학 예정자는 선호도를 바탕으로 일련의 기준에 의해 차등적으로 순위를 매기고, 이에 따라 지원할 용의를 가질 것이다.

### 가설2

이 순위에 대한 **보편적이고 정량적인** 평가 기준을 만들 수 있을 것이다.

### 3. 자료 수집

#### 자료 조사 대상

서울 소재 4년제 종합대학

(단, 서울교대, 한국체대 등 특수 대학 제외하여 총 29개 대학)

#### 수집 자료

(1) 2014년의 대학 선호도

(2) 2018년 대학별 공시 데이터

(3) 2020년도 대학별 공시 데이터

상관관계 분석    선호도에 영향 끼치는 지표에 대한 선형회귀모델 도출

대입

### 3. 자료 수집

순위	대학명	응답률
1	서울대학교	17.7
2	연세대학교	17.1
3	고려대학교	17.0
4	성균관대학교	15.1
5	한양대학교	13.2
6	중앙대학교	12.5
7	서강대학교	12.4
8	경희대학교	11.9
9	이화여자대학교	11.5
10	건국대학교	10.8
11	KAIST	10.0
12	서울시립대학교	9.4
13	동국대학교	8.4
14	홍익대학교	8.4
15	세종대학교	8.0
16	국민대학교	7.8
17	단국대학교	7.7
18	한국외국어대학교	7.6
19	송실대학교	5.9

1. 전국 고교생 1,000명을 대상으로 한 대학 선호도 설문조사 결과이며, 조사된 종합 대학은 본 조의 목표 조사 대상 대학보다 **큰 범위임**.
2. 2014년 고등학생의 **대학 선호도에 영향을 미치는 요인**은 2020년 현재까지 **동일하다고** 가정함.

▲ 2014년 내일신문 고등학생 대학 선호도 조사 결과

## 3. 자료 수집

### 2018년도 대학 데이터

대학명	기숙사	교지확보율	교사확보율	연구비	논문실적	전임교원확보율	장학금	교육비	외부경력교원	외국인교수	외국인학생	교관학생	취업률	유지취업률	중도탈락학생비율	창업활동	현장실습	창업교육	신입생충원	경장률	기회균형
서울대학교	21.60	730.30	304.00	229759.00	3103.50	132.56	3116800.00	48272588.90	0.67	107.00	197.00	1169.00	68.30	93.60	1.10	16.00	100.00	15.00	99.80	6.80	4.90
연세대학교	35.20	214.10	277.30	202906.60	2070.13	108.65	3242900.00	30241700.00	0.30	119.00	1177.00	3666.00	68.70	93.00	1.70	18.00	81.00	40.00	99.50	16.70	6.70
고려대학교	10.70	126.10	185.00	202769.94	1700.68	97.35	2985300.00	22859500.00	0.50	104.00	2327.00	2581.00	68.10	93.70	1.80	42.00	101.00	64.00	99.90	7.30	5.10
성균관대학교	21.70	106.30	187.70	217252.90	1890.77	124.17	3471700.00	28081400.00	0.66	112.00	274.00	1055.00	75.10	94.70	2.00	14.00	418.00	81.00	99.90	22.60	5.20
한양대학교	12.20	72.80	165.40	170695.98	1316.03	94.13	3295100.00	21388000.00	0.77	80.00	1608.00	3932.00	69.60	92.90	1.90	61.00	903.00	55.00	99.70	23.20	12.10
중앙대학교	11.60	28.80	129.60	108331.84	1293.48	95.39	2961600.00	15040500.00	0.22	42.00	1492.00	1319.00	67.80	92.10	2.70	16.00	696.00	93.00	99.80	26.00	13.70
서강대학교	12.10	76.30	173.70	138146.42	410.22	73.39	2805800.00	18676800.00	0.40	22.00	752.00	817.00	67.30	94.40	2.60	7.00	260.00	26.00	99.90	26.70	7.20
경희대학교	21.40	223.20	147.70	104443.21	1659.04	93.94	2996100.00	17232200.00	0.43	94.00	2770.00	2500.00	63.80	89.20	2.60	9.00	1552.00	112.00	99.70	19.00	11.50
이화여자대학교	21.40	104.10	188.40	107302.31	1127.73	9.82	3385500.00	19517100.00	0.52	61.00	969.00	1579.00	62.70	86.10	2.10	5.00	415.00	18.00	99.90	12.20	6.90
건국대학교	18.50	98.00	137.00	96278.38	908.76	77.74	3211600.00	16690800.00	0.36	67.00	1272.00	1218.00	60.90	90.60	2.70	25.00	813.00	55.00	99.50	16.80	9.80
서울시립대학교	8.00	161.10	134.60	96824.76	431.64	87.02	1645500.00	13627500.00	0.33	15.00	169.00	643.00	64.20	93.90	2.20	4.00	88.00	30.00	99.20	12.90	11.90
동국대학교	9.60	87.10	129.10	80368.98	792.90	94.20	3017600.00	14484200.00	0.38	131.00	1208.00	811.00	67.30	90.50	3.30	16.00	833.00	50.00	99.50	16.20	11.50
홍익대학교	12.90	28.00	152.40	41707.01	440.11	75.00	3812300.00	12024800.00	0.49	135.00	638.00	336.00	64.20	92.90	2.90	10.00	409.00	29.00	99.00	12.10	9.90
세종대학교	11.00	31.80	111.20	129446.40	485.59	84.94	3731800.00	15044100.00	0.57	69.00	813.00	965.00	64.90	88.60	3.50	23.00	325.00	32.00	99.55	12.50	7.80
국민대학교	11.30	45.10	121.10	82392.58	575.22	83.94	3152300.00	14273100.00	0.44	48.00	1765.00	1496.00	64.70	89.30	2.80	19.00	1265.00	125.00	99.50	9.50	11.40
단국대학교	19.10	126.30	136.50	45131.78	606.28	76.32	3214100.00	13425700.00	0.51	19.00	559.00	1108.00	62.90	89.80	3.10	8.00	248.00	135.00	99.70	16.00	10.90
한국외국어대학교	16.10	581.60	144.00	34641.33	891.39	90.99	3003500.00	12092000.00	0.32	187.00	1270.00	2628.00	60.10	88.60	2.90	2.00	638.00	85.00	99.20	11.60	6.80
순천대학교	11.80	34.80	107.90	57216.32	671.34	81.03	3289900.00	13568400.00	0.62	48.00	976.00	635.00	66.50	90.80	3.50	14.00	225.00	59.00	99.60	14.50	15.20

- 위 데이터는 2014년 대학 선호도 조사 대상에 대응하는 데이터임.
- 2014년과 가장 가까우면서 수집 가능한 데이터로서 2018년 데이터를 사용



## 3. 자료 수집

### 2020년도 대학 데이터

대학명	기숙사	교지확보율	교사확보율	연구비	교수당 외부 연구비	교수당 자체 연구비	논문실적	국내 학술지 논문	국제학술지 논문	전임교원확보율	장학금	교육비	외부경학교원	외국인교수	외국인학생	교원학생	취업률	유지취업률	중도탈락학생비율	창업활동	현장실습	창업교육	신입생충원	경쟁률	기회균형
상명대학교	10.30	98.80	112.90	86722.02	84884.01	1839.02	359.76	231.91	127.85	85.44	3258241.90	12726262.90	1.00	7.00	1114.00	416.00	64.30	82.80	3.80	7.00	186.00	21.00	99.00	10.00	10.30
서강대학교	12.80	82.50	173.00	120825.04	114327.29	6500.38	385.42	195.85	189.57	73.07	2674152.70	18024222.00	0.18	19.00	1176.00	944.00	70.40	95.10	3.10	5.00	300.00	29.00	99.50	23.20	8.60
서경대학교	5.00	29.20	138.70	28662.68	27491.09	474.92	191.01	164.02	26.99	86.59	3139615.20	11085138.80	0.53	47.00	596.00	77.00	70.20	86.70	5.40	6.00		18.00	99.80	19.80	9.10
서울여자대학교	12.10	115.70	128.70	28011.81	26528.70	2533.00	146.40	123.93	22.48	68.31	3418385.00	12678632.20	0.59	15.00	39.00	337.00	58.80	82.00	3.70	3.00	43.00	28.00	99.90	14.40	8.00
서울과학기술대학교	23.10	136.50	129.10	89612.19	83600.46	6011.74	419.71	186.31	233.40	72.83	3089574.00	14611666.70	0.44	4.00	470.00	754.00	66.70	87.40	3.20	31.00	624.00	82.00	100.00	12.20	15.50
서울대학교	21.80	730.10	306.50	236927.88	228138.11	8789.76	3058.85	806.04	2252.81	136.23	3180697.00	48272588.90	0.61	109.00	226.00	1207.00	70.10	92.60	1.30	13.00	55.00	12.00	70.10	6.30	4.80
서울시립대학교	11.40	161.10	153.00	95873.19	85790.56	10082.63	444.78	295.19	149.59	88.48	1664083.00	14556763.00	0.33	6.00	252.00	899.00	68.50	93.30	2.50	3.00	121.00	35.00	68.50	13.30	13.00
건국대학교	18.50	94.70	129.60	110939.33	102471.90	8467.43	926.07	450.11	475.96	74.89	3162016.30	15990918.20	0.46	50.00	1372.00	1401.00	64.50	90.70	2.70	30.00	971.00	92.00	99.30	19.60	10.80
경희대학교	21.40	221.90	146.70	99872.72	94866.61	5006.09	1611.37	686.88	924.49	95.33	3081884.70	16480634.60	0.51	40.00	2939.00	2319.00	68.30	89.70	2.80	29.00	1758.00	163.00	99.50	16.60	12.20
고려대학교	10.70	126.00	186.10	243223.80	226072.20	17151.60	1797.62	557.55	1240.07	99.86	3211269.60	24559193.40	0.39	86.00	2154.00	2387.00	70.30	93.70	2.20	34.00	375.00	40.00	99.80	8.00	6.20
광운대학교	11.70	30.00	119.90	91379.08	86261.09	5117.98	355.80	181.39	174.41	78.80	2975822.70	14327089.30	0.46	25.00	733.00	207.00	64.90	91.40	3.70	12.00	307.00	22.00	99.90	13.30	15.00
국민대학교	11.60	45.40	121.80	72766.95	72396.58	370.37	616.72	429.79	186.93	90.85	3282095.80	14758081.80	0.50	49.00	1839.00	1403.00	67.00	87.50	2.70	23.00	984.00	138.00	99.80	9.20	12.30
덕성여자대학교	13.10	123.10	139.80	38985.73	35889.57	3096.15	164.99	130.57	34.42	73.39	2997055.70	12825624.10	0.27	19.00	25.00	227.00	62.00	81.20	3.80	1.00	210.00	12.00	99.70	12.50	5.40
동국대학교	10.10	87.80	130.00	93216.40	89822.52	3393.88	844.80	448.41	396.39	89.97	3223415.90	14739781.60	0.41	116.00	1248.00	832.00	67.40	92.30	3.00	12.00	679.00	62.00	99.80	17.70	12.50
동덕여자대학교	11.10	28.30	109.10	14496.12	12998.32	1497.80	163.80	141.66	22.14	70.06	3194239.60	11079090.00	0.58	17.00	27.00	190.00	62.40	82.80	4.30	5.00	166.00	8.00	99.50	13.60	9.30
명지대학교	29.10	209.60	146.00	70872.63	70650.03	222.60	138.66	42.72	95.94	82.72	3520825.40	13792188.80	0.48	19.00	181.00	148.00	69.60	86.70	4.00	0.00	294.00	16.00	99.90	9.00	7.70
삼육대학교	23.00	198.90	131.50	23551.59	21784.92	1766.67	142.76	113.51	29.25	83.00	3937121.60	13280473.30	0.64	31.00	111.00	112.00	67.20	80.40	4.30	16.00	199.00	27.00	99.90	12.70	7.00
성균관대학교	21.30	106.20	188.50	245294.35	224329.23	20965.10	2012.78	505.62	1507.17	129.42	3388267.30	27885103.10	0.61	105.00	2677.00	1433.00	77.00	95.40	2.30	15.00	752.00	158.00	99.70	19.40	5.70
성신여자대학교	9.80	56.30	135.30	19242.51	15876.49	3366.01	271.53	231.61	39.93	66.67	3815886.70	12623320.10	0.44	25.00	22.00	724.00	59.60	77.90	3.30	8.00	190.00	31.00	100.00	11.10	7.20
세종대학교	11.20	31.60	134.70	129079.76	123738.35	3332.41	557.54	176.22	381.32	84.18	3184328.70	15857592.70	0.49	71.00	924.00	845.00	64.20	85.70	5.50	27.00	314.00	20.00	100.00	15.80	9.60
숙명여자대학교	11.80	22.50	126.40	48740.26	43899.30	2686.13	415.64	287.18	128.46	78.22	3347472.50	14491687.90	0.57	30.00	68.00	799.00	64.40	86.90	2.80	10.00	312.00	43.00	99.60	11.70	12.00
숭실대학교	12.00	34.70	107.80	65404.85	61582.30	3822.54	653.81	449.08	204.73	81.34	3013314.10	13224918.20	0.54	37.00	1349.00	760.00	69.70	91.10	2.70	8.00	321.00	108.00	99.90	12.90	16.20
연세대학교	36.70	218.50	293.50	226428.48	219985.34	6443.14	2155.40	593.75	1561.65	114.25	3419692.60	30591101.00	0.50	129.00	1279.00	4741.00	70.10	91.30	2.00	24.00	84.00	57.00	99.80	13.80	6.70
이화여자대학교	22.20	103.80	192.80	118767.97	107464.09	11306.97	1035.14	582.69	452.46	90.93	3263233.90	19075139.00	0.66	56.00	975.00	1507.00	62.10	88.40	2.30	10.00	321.00	22.00	99.80	9.60	6.90
중앙대학교	12.30	29.80	134.30	126540.46	115850.98	10689.48	1243.66	610.72	632.95	95.64	3042138.60	17223235.10	0.53	40.00	1815.00	1969.00	69.70	91.30	2.50	9.00	760.00	79.00	99.00	24.40	14.80
한국외국어대학교	17.90	582.90	146.60	28394.58	23528.43	4866.15	792.04	561.96	230.08	88.54	2561252.10	11345010.90	0.35	182.00	1776.00	2675.00	63.70	90.20	3.50	23.00	498.00	100.00	99.60	11.20	9.60
한성대학교	11.20	35.10	123.60	23724.04	20227.60	3506.37	190.07	137.44	52.65	82.13	3393165.10	12496919.90	0.60	15.00	182.00	207.00	64.70	86.00	4.80	15.00	241.00	40.00	99.50	11.60	10.50
한양대학교	11.70	73.60	164.50	177671.47	166263.80	11407.67	1249.45	455.31	794.15	96.90	3394820.40	21357041.20	0.62	73.00	1856.00	3849.00	73.40	91.70	2.60	43.00	843.00	512.00	99.50	21.00	12.30
홍익대학교	12.80	28.20	155.80	46282.16	37725.15	8557.02	457.18	314.59	142.60	79.52	3896006.50	12391304.40	0.70	96.00	1092.00	413.00	68.20	93.90	3.60	25.00	420.00	81.00	100.00	14.10	9.80

앞선 데이터(2018년도 자료)로 선형회귀모델 도출 뒤 2020년 대학 데이터 대입

### 3. 자료 수집

#### 수집한 대학교 데이터의 구성 항목 (총 21개)

기숙사	대학의 기숙사 수용률	교환학생	외국대학과의 학점교류 현황
교지확보율	대학의 교지확보율	취업률	졸업생 취업률
교사확보율	대학의 교사시설 확보율	유지취업률	졸업생 유지취업률(2차 기준)
연구비	전임교원 1인당 평균 연구비 지원	중도탈락학생비율	전체학생 중 중도탈락 학생의 비율
논문실적	전임교원 1인당 평균 논문실적	창업활동	졸업생 창업 현황
전임교원확보율	전임교원 확보 비율	현장실습	대학의 학생 현장실습 참여 학생 현황
교육비	대학의 학생 1인당 평균 교육비 투자	창업교육	대학의 창업교육 지원 현황
장학금	재학생 1인당 평균 장학금	신입생충원	신입생 충원률
외국인 교수	외국인 교수 보유 현황	경쟁률	대학 입학 경쟁률
외부경력교원	최근 5년간 산업체 경력 교원 고용 비율	기회균형	기회균형 전형 선발 비율
외국인학생	학위과정 외국인 학생 등록 현황		

## 4. 상관관계 분석

### 상관관계와 상관계수

#### 상관관계

변수들 사이의 선형관계를 나타낸 것

한 변수의 변화와 다른 변수의 변화가 서로 어떤 패턴을 가지는 것을 의미

#### 상관계수

두 변수의 “관계성”과 “선형성”의 척도

$$r = \frac{\sum_i (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_i (X_i - \bar{X})^2} \sqrt{\sum_i (Y_i - \bar{Y})^2}}$$

## 4. 상관관계 분석

### 상관관계와 상관계수

#### 상관계수의 해석

- 양(+)의 값을 가질 때는 **정의 상관관계**를 가진다.
- 음(-)의 값을 가질 때는 **부의 상관관계**를 가진다.
- 1 또는 -1에 가까워질수록 강한 상관관계를 갖는다.
- 0일 때는 변수 간에 **아무 관련성이 없음**을 의미한다.

범위	선형관계
$-1.0 \leq r \leq -0.7$	강한 음의 상관관계
$-0.7 < r \leq -0.3$	음의 상관관계
$-0.3 < r \leq -0.1$	약한 음의 상관관계
$-0.1 < r < +0.1$	상관관계 거의 없음
$+0.1 \leq r < +0.3$	약한 양적 상관관계
$+0.3 \leq r < +0.7$	양적 상관관계
$+0.7 \leq r \leq +1.0$	강한 양적 상관관계

#### 분석 목적

대학의 어떤 요소가 대학 선호도에 영향을 끼치는지 규명하기 위해 상관관계 분석을 실시함

## 4. 상관관계 분석

### 각 항목별 상관계수 도출

#### 상관관계 분석 실행

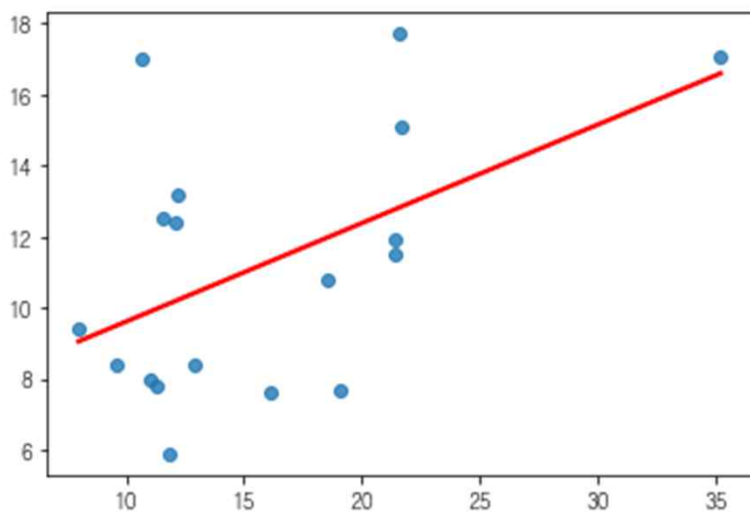
- 상관계수 계산 수단으로 프로그래밍 언어 파이썬(Python) 사용
- $Y =$  [17.7, 17.1, 17.0, ..... , 5.9]
- $X =$  수집한 데이터 항목들 [ '대학의 기숙사 수용률' , '대학의 졸업생 취업률' ,  
'대학의 학생 1인당 평균 장학금 지급액' , .....]

순위	대학명	응답률
1	서울대학교	17.7
2	연세대학교	17.1
3	고려대학교	17.0
4	성균관대학교	15.1
5	한양대학교	13.2
6	중앙대학교	12.5
7	서강대학교	12.4
8	경희대학교	11.9
9	이화여자대학교	11.5
10	건국대학교	10.8
11	KAIST	10.0
12	서울시립대학교	9.4
13	동국대학교	8.4
14	홍익대학교	8.4
15	세종대학교	8.0
16	국민대학교	7.8
17	단국대학교	7.7
18	한국외국어대학교	7.6
19	송실대학교	5.9

## 4. 상관관계 분석

### 각 항목별 상관계수 도출

#### 1. 대학의 기숙사 수용률



X축 : 대학의 학생정원 기준 기숙사 수용률

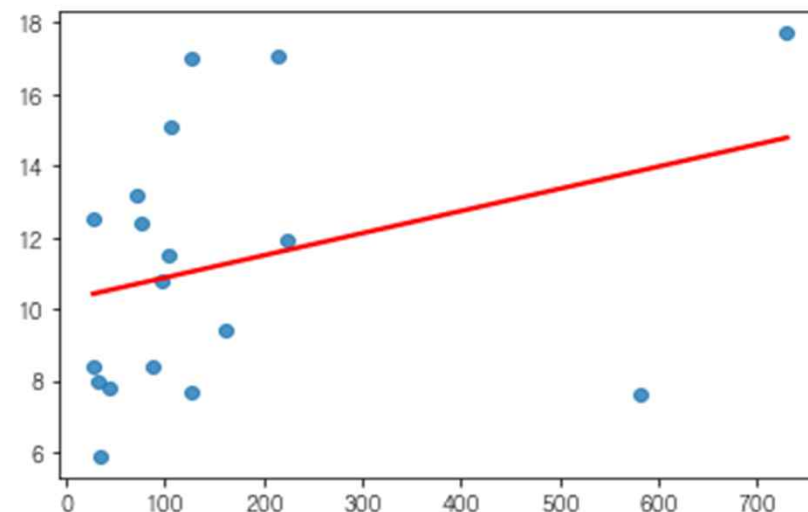
Y축 : 대학 선호도 조사 응답률

\* 상관계수 : 0.508

대학의 기숙사 수용률과 고등학생의 대학 선호도

응답률은 양의 상관관계를 갖는다.

#### 2. 대학의 교지(校地) 확보율



X축 : 대학의 학생정원 기준 교지(校地) 확보율

Y축 : 대학 선호도 조사 응답률

\* 상관계수 : 0.325

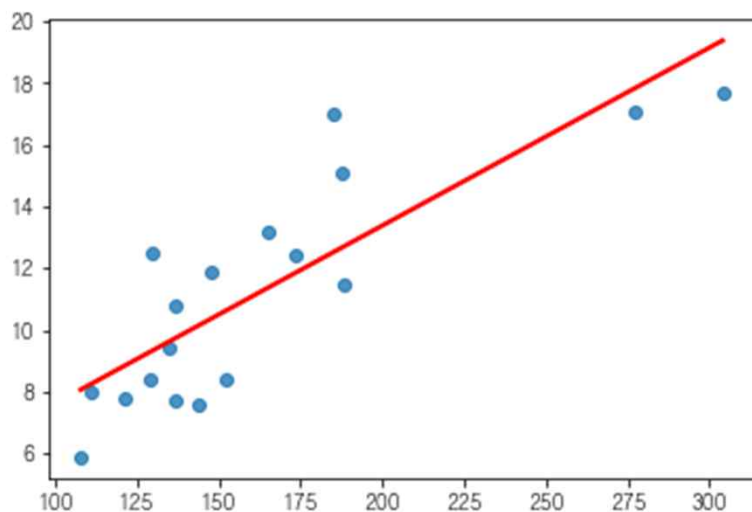
대학의 교지(校地) 확보율과 고등학생의 대학 선호도

응답률은 양의 상관관계를 갖는다.

## 4. 상관관계 분석

### 각 항목별 상관계수 도출

#### 3. 대학의 교사(校舍)시설 확보율



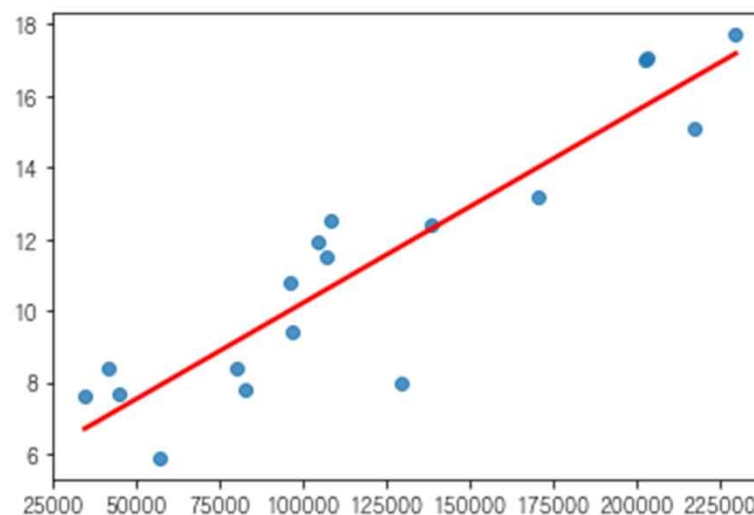
X축 : 대학의 학생정원 기준 교사(校舍)시설 확보율

Y축 : 대학 선호도 조사 응답률

\* 상관계수 : 0.837

대학의 교사(校舍)시설 확보율과 고등학생의 대학 선호도  
응답률은 **강한 양의 상관관계**를 갖는다.

#### 4. 대학의 전임교원 1인당 평균 연구비 지원



X축 : 대학의 전임교원 1인당 평균 연구비 지원액

Y축 : 대학 선호도 조사 응답률

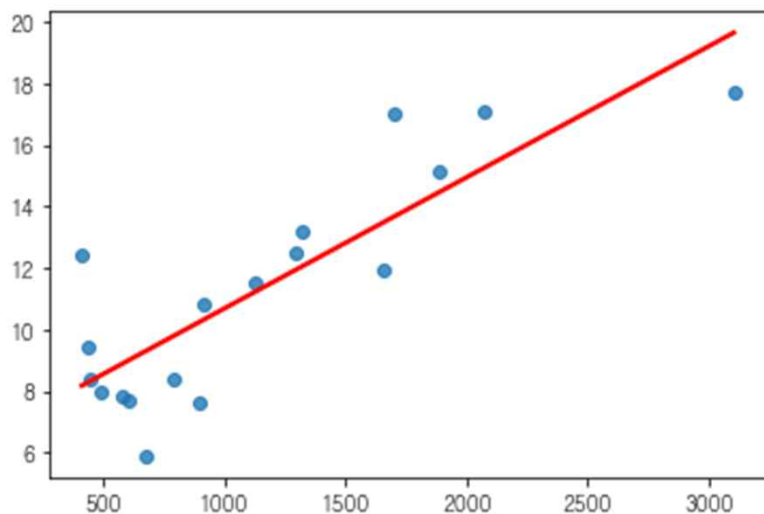
\* 상관계수 : 0.914

대학의 전임교원 1인당 평균 연구비 지원과 고등학생의  
대학 선호도 응답률은 **강한 양의 상관관계**를 갖는다.

## 4. 상관관계 분석

### 각 항목별 상관계수 도출

#### 5. 대학의 전임교원 1인당 평균 논문실적



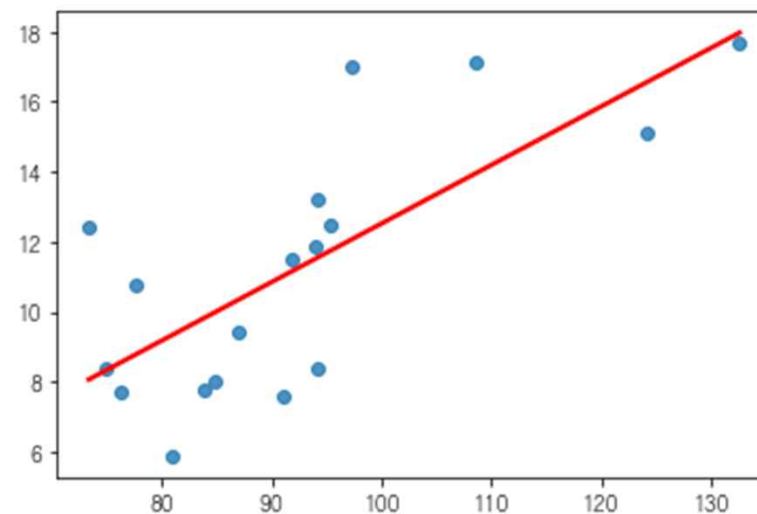
X축 : 대학의 전임교원 1인당 평균 논문실적

Y축 : 대학 선호도 조사 응답률

\* 상관계수 : 0.843

대학의 전임교원 1인당 평균 논문실적과 고등학생의  
대학 선호도 응답률은 **강한 양의 상관관계**를 갖는다.

#### 6. 대학의 전임교원 확보율



X축 : 대학의 학생정원 기준 전임교원 확보율

Y축 : 대학 선호도 조사 응답률

\* 상관계수 : 0.731

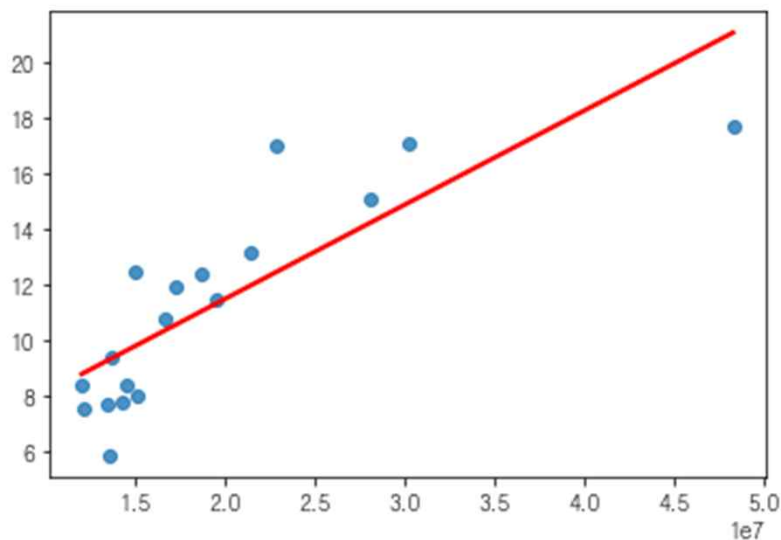
대학의 전임교원 확보율과 고등학생의 대학 선호도  
응답률은 **강한 양의 상관관계**를 갖는다.



## 4. 상관관계 분석

### 각 항목별 상관계수 도출

#### 7. 대학의 학생 1인당 평균 교육비 투자액

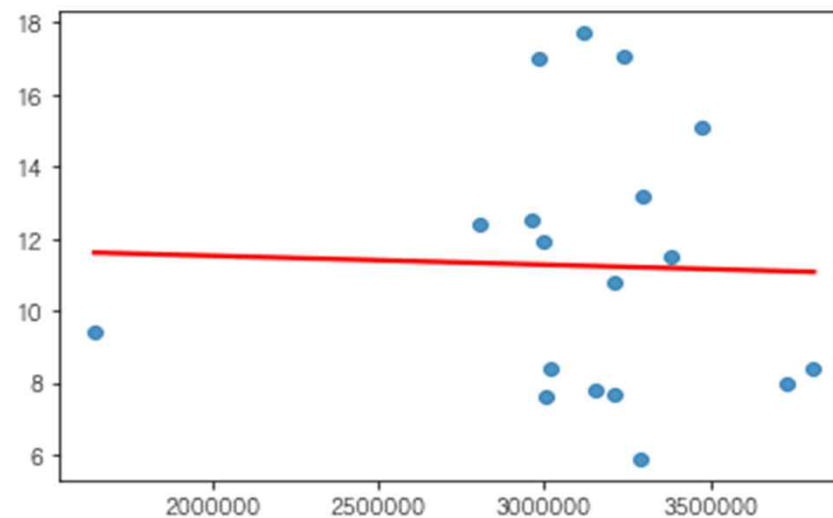


X축 : 대학의 학생 1인당 평균 교육비 투자액  
Y축 : 대학 선호도 조사 응답률

\* 상관계수 : 0.823

대학의 전임교원 1인당 평균 논문실적과 고등학생의  
대학 선호도 응답률은 **강한 양의 상관관계**를 갖는다.

#### 8. 대학의 재학생 1인당 평균 장학금 지급액



X축 : 대학의 재학생 1인당 평균 장학금 지급액  
Y축 : 대학 선호도 조사 응답률

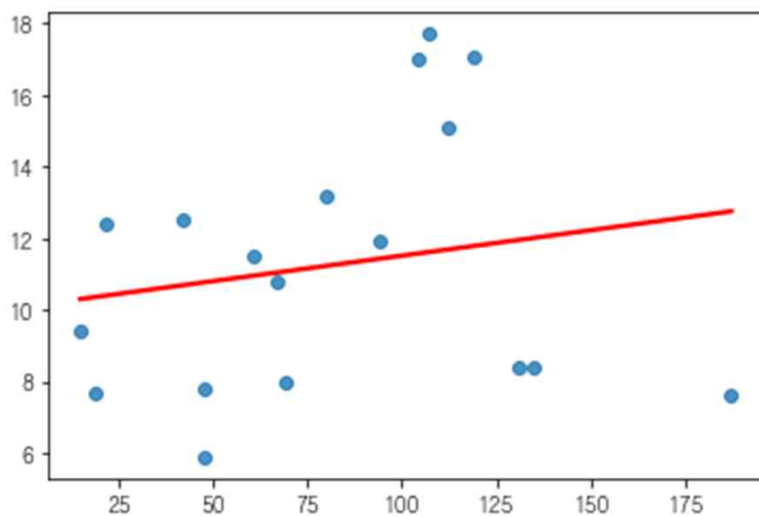
\* 상관계수 : -0.031

대학의 재학생 1인당 평균 장학금 지급액과 고등학생의  
대학 선호도 응답률은 **상관관계가 거의 없다**.

## 4. 상관관계 분석

### 각 항목별 상관계수 도출

#### 9. 대학의 외국인 교수 보유 현황



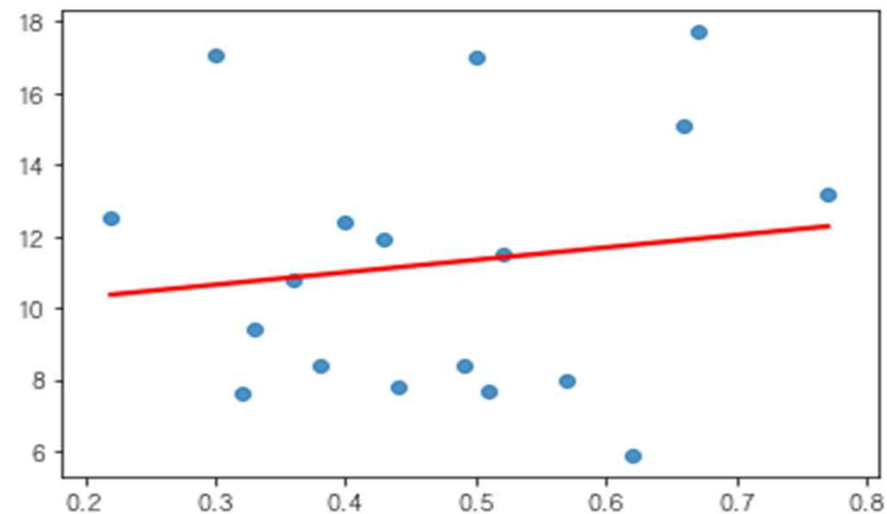
X축 : 대학의 외국인 교수 보유 수

Y축 : 대학 선호도 조사 응답률

\* 상관계수 : 0.181

대학의 외국인 교수 보유 현황과 고등학생의 대학 선호도 응답률은 **약한 양의 상관관계**를 갖는다.

#### 10. 대학의 최근 5년간 산업체 경력 교원 고용 비율



X축 : 대학의 최근 5년간 산업체 경력 교원 고용 비율

Y축 : 대학 선호도 조사 응답률

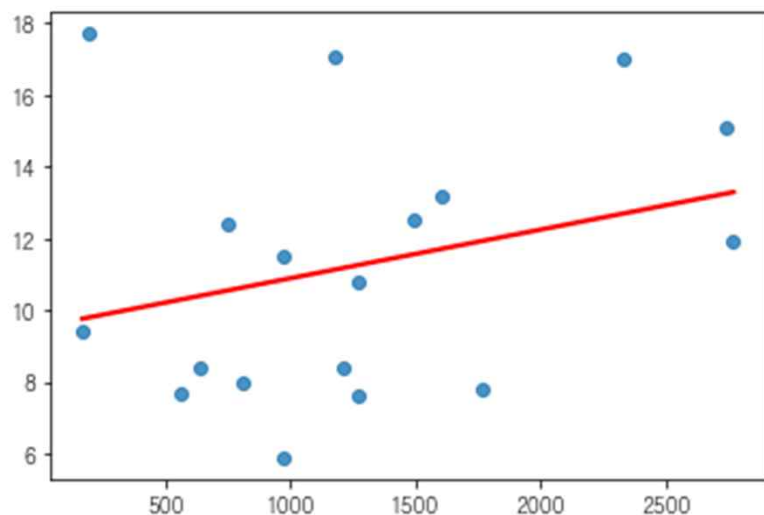
\* 상관계수 : 0.139

대학의 최근 5년간 산업체 경력 교원 고용 비율과 고등학생의 대학 선호도 응답률은 **약한 양의 상관관계**를 갖는다.

## 4. 상관관계 분석

### 각 항목별 상관계수 도출

#### 11. 대학의 학위과정 외국인학생 등록 현황

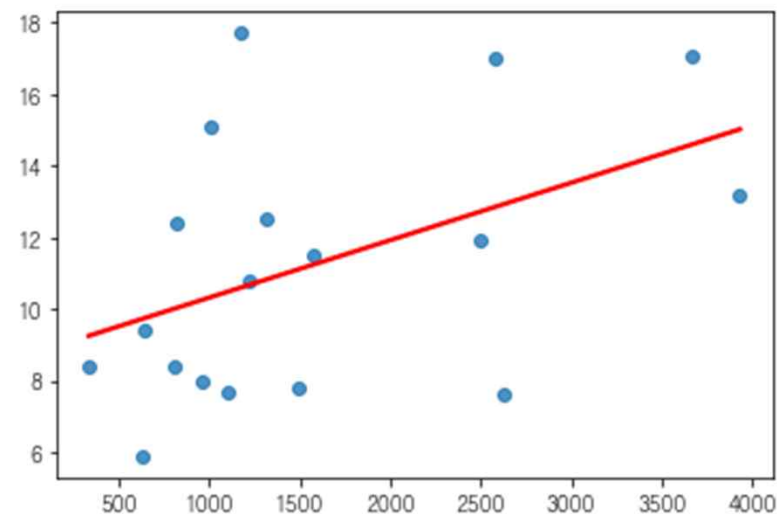


X축 : 대학의 학위과정 외국인학생 등록 수  
Y축 : 대학 선호도 조사 응답률

\* 상관계수 : 0.283

대학의 학위과정 외국인학생 등록 현황과 고등학생의  
대학 선호도 응답률은 **약한 양의 상관관계**를 갖는다.

#### 12. 대학의 외국 대학과의 학점교류 현황



X축 : 대학의 교환학생 수 (파견인원 + 유치인원)  
Y축 : 대학 선호도 조사 응답률

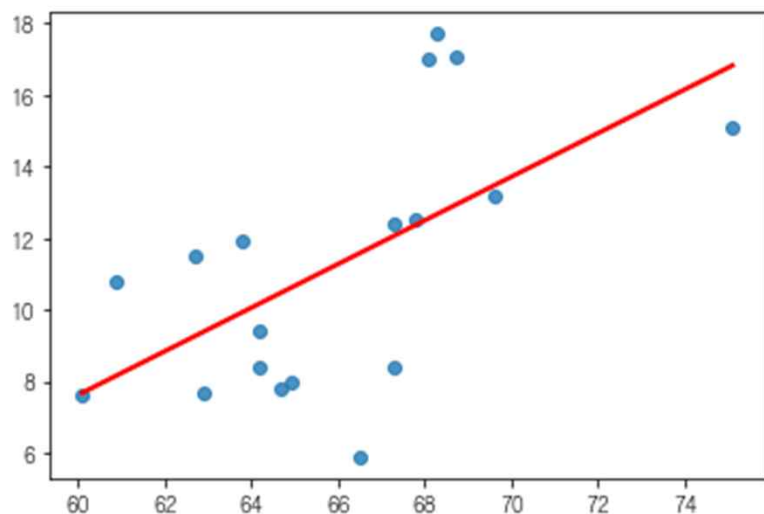
\* 상관계수 : 0.459

대학의 외국 대학과의 학점교류 현황과 고등학생의  
대학 선호도 응답률은 **양의 상관관계**를 갖는다.

## 4. 상관관계 분석

### 각 항목별 상관계수 도출

#### 13. 대학의 졸업생 취업률

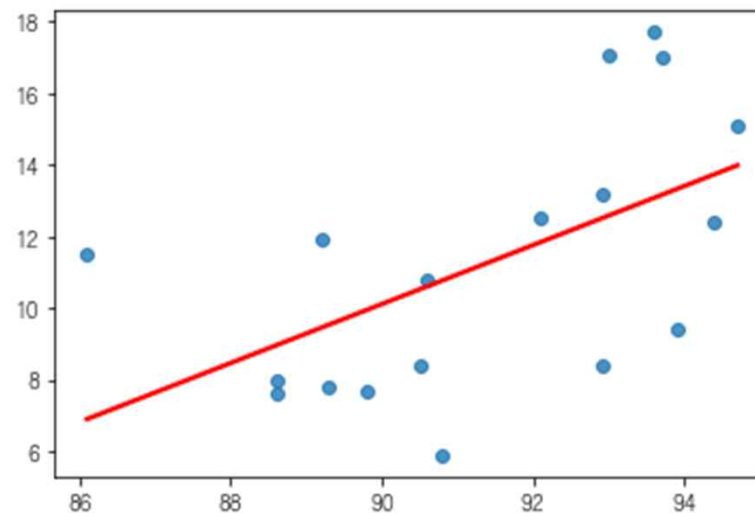


X축 : 대학의 졸업생 취업률  
Y축 : 대학 선호도 조사 응답률

★ 상관계수 : 0.593

대학의 졸업생 취업률과 고등학생의 대학 선호도  
응답률은 **양의 상관관계**를 갖는다.

#### 14. 대학의 졸업생 유지 취업률(2차)



X축 : 대학의 졸업생 유지 취업률(2차)  
Y축 : 대학 선호도 조사 응답률

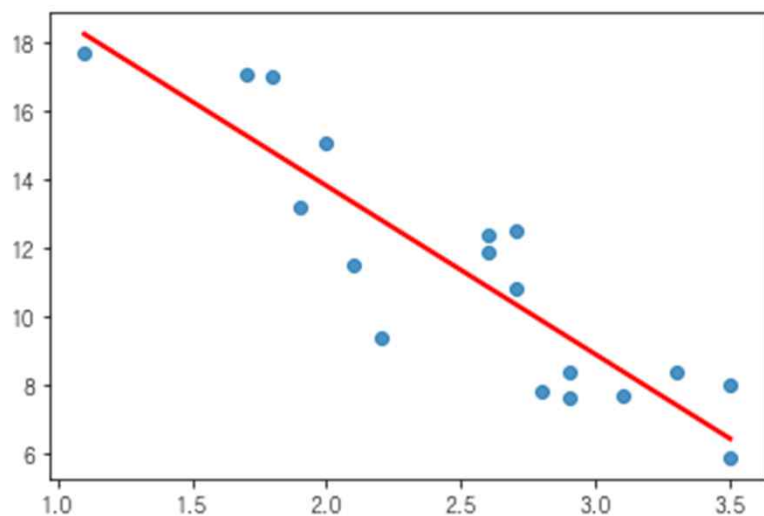
★ 상관계수 : 0.550

대학의 졸업생 유지 취업률과 고등학생의 대학 선호도  
응답률은 **양의 상관관계**를 갖는다.

## 4. 상관관계 분석

### 각 항목별 상관계수 도출

#### 15. 대학의 중도탈락 학생 비율

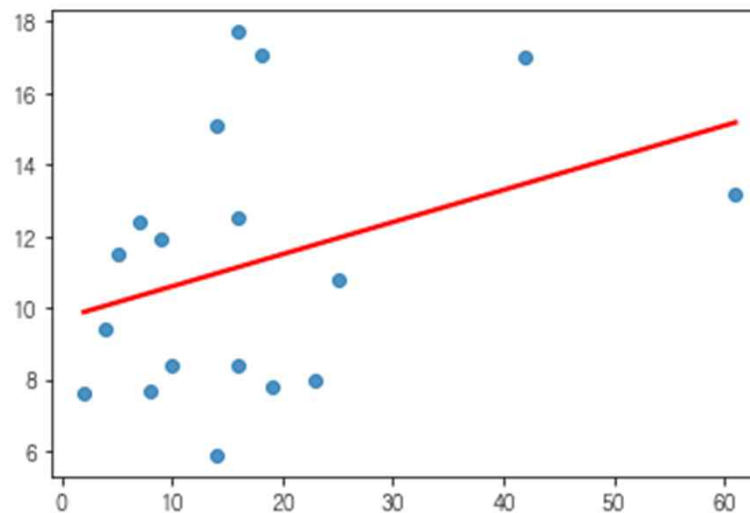


X축 : 대학의 중도탈락 학생 비율  
Y축 : 대학 선호도 조사 응답률

\* 상관계수 :  $-0.890$

대학의 중도탈락 학생 비율과 고등학생의 대학 선호도 응답률은  
강한 음의 상관관계를 갖는다.

#### 16. 대학의 졸업생 창업 현황



X축 : 대학의 졸업생 창업자 수  
Y축 : 대학 선호도 조사 응답률

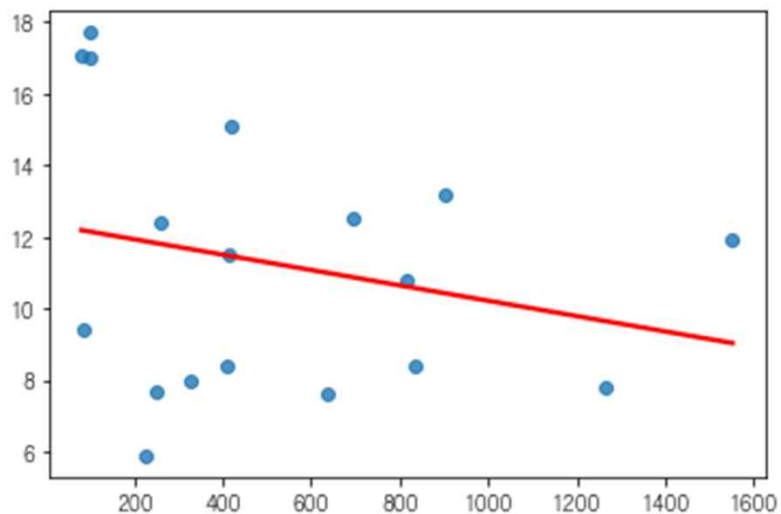
\* 상관계수 :  $0.353$

대학의 졸업생 창업현황과 고등학생의 대학 선호도  
응답률은 양의 상관관계를 갖는다.

## 4. 상관관계 분석

### 각 항목별 상관계수 도출

#### 17. 대학의 학생 현장실습 참여 현황

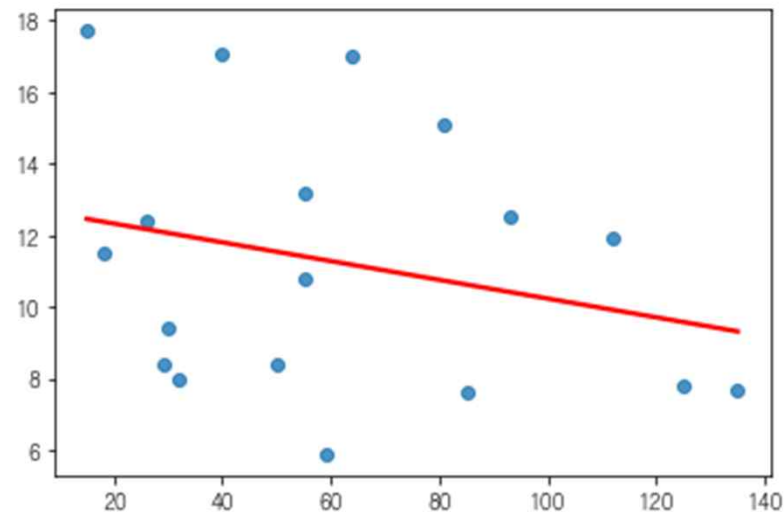


X축 : 대학의 현장실습 프로그램 참여 학생 총합  
Y축 : 대학 선호도 조사 응답률

\* 상관계수 :  $-0.247$

대학의 중도탈락 학생 비율과 고등학생의 대학 선호도  
응답률은 **약한 음의 상관관계**를 갖는다.

#### 18. 대학의 창업교육 지원 현황



X축 : 대학의 창업교육 강좌 개설 총합  
Y축 : 대학 선호도 조사 응답률

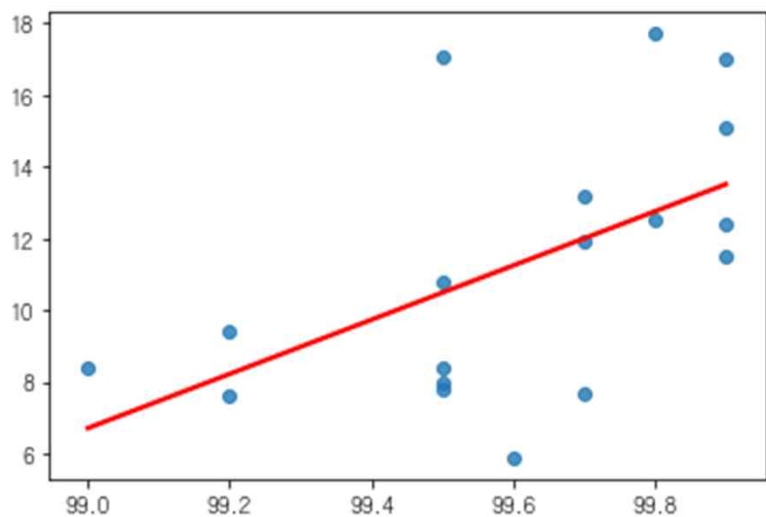
\* 상관계수 :  $-0.262$

대학의 창업교육 지원 현황과 고등학생의 대학 선호도  
응답률은 **약한 음의 상관관계**를 갖는다.

## 4. 상관관계 분석

### 각 항목별 상관계수 도출

#### 19. 대학의 신입생 충원률

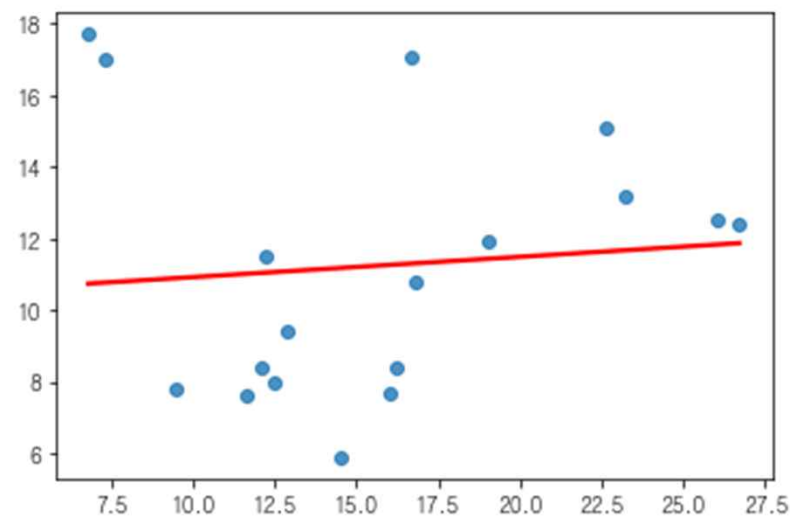


X축 : 대학의 신입생 충원률  
Y축 : 대학 선호도 조사 응답률

\* 상관계수 : 0.547

대학의 신입생 충원률과 고등학생의 대학 선호도  
응답률은 **양의 상관관계**를 갖는다.

#### 20. 대학의 입학 경쟁률



X축 : 대학의 입학 경쟁률  
Y축 : 대학 선호도 조사 응답률

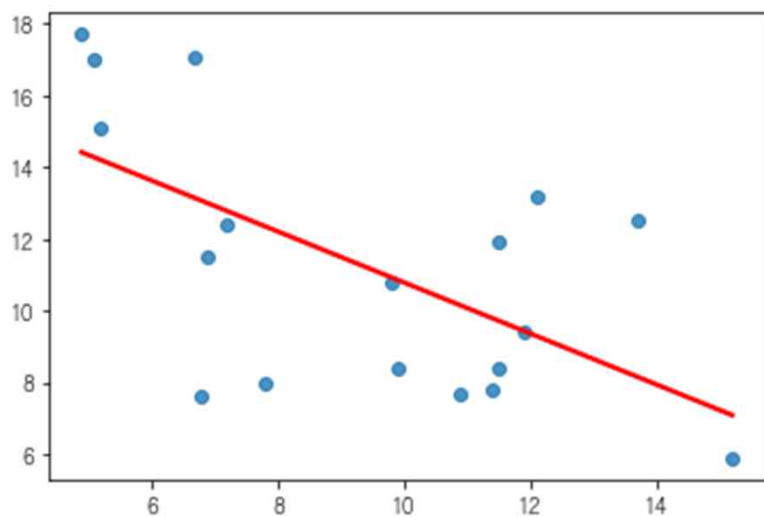
\* 상관계수 : 0.092

대학의 입학 경쟁률과 고등학생의 대학 선호도 응답률은  
**상관관계가 거의 없다.**

## 4. 상관관계 분석

### 각 항목별 상관계수 도출

#### 21. 대학의 기회균형 전형 선발비율



X축 : 대학의 기회균형 전형 선발비율

Y축 : 대학 선호도 조사 응답률

\* 상관계수 :  $-0.610$

대학의 기회균형 전형 선발비율과 고등학생의 대학 선호도  
응답률은 **음의 상관관계**를 갖는다.



## 4. 상관관계 분석

### 각 항목별 상관계수 도출

결과에 따른 이후 조사에 사용할 항목 분류

변수	상관계수
기숙사 수용률	0.508
교지확보율	0.325
교사시설 확보율	0.832
전임교원 1인당 평균 연구비 지원	0.914
전임교원 1인당 평균 논문 실적	0.843
학생정원 기준 전임교원 확보율	0.732
학생 1인당 평균 장학금 지급	-0.031
학생 1인당 평균 교육비	0.823
최근 5년간 외부경력 교원 임용 비율	0.139
외국인 교수 고용 현황	0.181
학사학위 과정 외국인학생 현황	0.283

변수	상관계수
외국대학과의 학점교류 현황	0.459
대학의 졸업생 취업률	0.593
졸업생 유지취업률(2차)	0.550
중도탈락 학생 비율	-0.890
졸업생의 창업활동	0.353
현장실습 참여 현황	-0.247
창업교육 지원 현황	-0.262
신입생 충원률	0.549
입학 경쟁률	0.092
기회균형 선발 비율	-0.607

상관관계가 거의 없는 ‘장학금’ 과 ‘경쟁률’ 항목을 제외하고, 나머지 항목들을 이후 분석에 이용

## 5. 선형회귀분석

### 다중선형회귀분석과 최소제곱법

#### 다중선형회귀분석

- 선형회귀분석 중에서도, **두 개 이상의 독립변수**가 **동시에** 종속변수에 미치는 영향을 분석하는 방법
- 종속변수를 가장 잘 설명할 수 있는 **최선의 회귀직선**을 찾는 것을 분석 목표로 함
- 다중선형회귀에서의 추정회귀직선:  $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 \dots \dots \hat{\beta}_n x_n$   
 $(\hat{y}: \quad, x_n: \quad, \hat{\beta}_n: \quad, \hat{\beta}_0: \quad)$
- 최선의 회귀직선 도출을 위해 최소제곱법을 사용

#### 최소제곱법

- 잔차( $\epsilon = y - \hat{y}$ )의 합이 **최소**가 되도록 하는 회귀직선을 찾는 방법
- 잔차 : 실제 종속변수의 값과 추정회귀직선에서의 종속변수 값의 차이

## 5. 선형회귀분석

### 다중선형회귀분석과 최소제곱법

#### 분석 목적

- (1) 고등학생의 대학 선호도 응답률( $y$ )을 가장 잘 설명할 수 있는 회귀직선 ( $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1x_1 + \hat{\beta}_2x_2 \dots \hat{\beta}_nx_n$ ) 을 도출하고,
- (2) 이를 바탕으로 2020년 서울 소재 종합 대학의 선호도 응답률을 예측하고자 함
- (3) 예측한 선호도 응답률을 바탕으로 새로운 대학 순위를 제시하고자 함

#### 독립변수 설정

상관관계 분석의 결과로 상관관계가 없다고 판정된 ‘경쟁률’ 항목과 ‘장학금’ 항목을 제외한 19개의 변수를 모두 독립변수로 사용함

## 5-1. 데이터 정규화

### 사용 이유

- 현재 데이터의 각 항목들은 비율, 금액, 사람수 등 다양한 단위로 구성됨
- 또한, 단위가 금액인 항목의 경우 절대적인 값의 크기가 다른 값들에 비해 크므로 선형화귀분석 시 회귀식이 편향될 수 있음
- 따라서, 항목 간 비교와 선형화귀분석의 정확도를 위해 각 항목들의 값의 단위와 범위를 일치시킬 필요가 있음

### 전제 조건

1. 값의 범위를 0 ~ 1 사이로 고정함.
2. 정규화 이후에도 데이터의 분포는 전과 동일 -> 값의 범위만 변경됨
3. MIN MAX 정규화 사용

## 5-1. 데이터 정규화

	기숙사	교지확보 율	교사확보 율	연구비	논문실적	전임교원확보 율	교육비	외부경력교 원	외국인교 수	외국인학 생	교환학생	취업률	유지취업 률	중도탈락학생비 율	창업활동	현장실습	창업교육	신입생충 원	기회균형
0	0.500000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	0.818182	0.534884	0.010765	0.231646	0.546667	0.872093	0.000000	0.237288	0.012916	0.000000	0.888889	0.000000
1	1.000000	0.264986	0.863845	0.862378	0.616315	0.595910	0.502566	0.145455	0.604651	0.387543	0.926029	0.573333	0.802326	0.250000	0.271186	0.000000	0.208333	0.555556	0.174757
2	0.099265	0.139684	0.393167	0.861678	0.479141	0.404935	0.298907	0.509091	0.517442	0.829681	0.624305	0.533333	0.883721	0.291667	0.677966	0.013596	0.408333	1.000000	0.019417
3	0.503676	0.111491	0.406935	0.935905	0.549720	0.858205	0.442968	0.800000	0.563953	0.989235	0.186040	1.000000	1.000000	0.375000	0.203390	0.229096	0.550000	1.000000	0.029126
4	0.154412	0.063790	0.293218	0.697295	0.336322	0.350515	0.258311	1.000000	0.377907	0.553249	1.000000	0.633333	0.790698	0.333333	1.000000	0.558804	0.333333	0.777778	0.699029
5	0.132353	0.001139	0.110658	0.377672	0.327950	0.371810	0.083197	0.000000	0.156977	0.508651	0.273359	0.513333	0.697674	0.666667	0.237288	0.418083	0.650000	0.888889	0.854369
6	0.150735	0.068774	0.335543	0.530475	0.000000	0.000000	0.183515	0.327273	0.040698	0.224145	0.133760	0.480000	0.965116	0.625000	0.084746	0.121686	0.091667	1.000000	0.223301
7	0.492647	0.277944	0.202958	0.357742	0.463680	0.347304	0.143661	0.381818	0.459302	1.000000	0.601780	0.246667	0.360465	0.625000	0.118644	1.000000	0.808333	0.777778	0.640777
8	0.492647	0.108358	0.410505	0.372396	0.266408	0.311475	0.206697	0.545455	0.267442	0.307574	0.345662	0.173333	0.000000	0.416667	0.050847	0.227056	0.025000	1.000000	0.194175
9	0.386029	0.099673	0.148394	0.315897	0.185105	0.073517	0.128725	0.254545	0.302326	0.424068	0.245273	0.053333	0.523256	0.666667	0.389831	0.497621	0.333333	0.555556	0.475728
10	0.000000	0.189520	0.136155	0.318697	0.007953	0.230353	0.044215	0.200000	0.000000	0.000000	0.085373	0.273333	0.906977	0.458333	0.033898	0.004759	0.125000	0.222222	0.679612
11	0.058824	0.084152	0.108108	0.234359	0.142087	0.351698	0.067850	0.290909	0.674419	0.399462	0.132091	0.480000	0.511628	0.916667	0.237288	0.511217	0.291667	0.555556	0.640777
12	0.180147	0.000000	0.226925	0.036212	0.011098	0.027210	0.000000	0.490909	0.697674	0.180315	0.000000	0.273333	0.790698	0.750000	0.135593	0.222978	0.116667	0.000000	0.485437
13	0.110294	0.005411	0.016828	0.485887	0.027984	0.195200	0.083296	0.636364	0.313953	0.247597	0.174917	0.320000	0.290698	1.000000	0.355932	0.165874	0.141667	0.555556	0.281553
14	0.121324	0.024349	0.067313	0.244731	0.061264	0.178300	0.062026	0.400000	0.191860	0.613610	0.322581	0.306667	0.372093	0.708333	0.288136	0.804895	0.916667	0.555556	0.631068
15	0.408088	0.139969	0.145844	0.053765	0.072796	0.049518	0.038648	0.527273	0.023256	0.149942	0.214683	0.186667	0.430233	0.833333	0.101695	0.113528	1.000000	0.777778	0.582524
16	0.297794	0.788267	0.184090	0.000000	0.178656	0.297448	0.001854	0.181818	1.000000	0.423299	0.637375	0.000000	0.290698	0.750000	0.000000	0.378654	0.583333	0.222222	0.184466
17	0.139706	0.009682	0.000000	0.115699	0.096952	0.129119	0.042585	0.727273	0.191860	0.310265	0.083148	0.426667	0.546512	1.000000	0.203390	0.097893	0.366667	0.666667	1.000000

▲ 항목들을 정규화한 결과를 나타낸 표

## 5-2. 회귀분석 결과

### 다중선형회귀분석 결과

파이썬 코드를 통해 회귀분석 진행

#### ① 회귀식의 기울기

```
In [33]: model1.coef_
```

```
Out[33]: array([-0.91725493, -4.67228452, -0.78279637,  2.2653429, 13.50522401,
               -1.82347648, -4.68199861, -1.32967242,  0.04130948, -3.23905999,
                2.29747415, -1.25637984,  4.25914509, -2.2546124 , -1.29384865,
                1.97365249, -0.57078926,  2.10366704, -4.25139286])
```

#### ② 회귀식의 절편

```
In [34]: model1.intercept_
```

```
Out[34]: 10.30095710472393
```

#### ③ 도출된 회귀식

$$\begin{aligned} \hat{y} = & -0.91725493x_1 - 4.67228452x_2 - 0.78279637x_3 + 2.2653429x_4 + 13.50522401x_5 \\ & - 1.82347648x_6 - 4.68199861x_7 - 1.32967242x_8 + 0.04130948x_9 - 3.23905999x_{10} \\ & + 2.29747415x_{11} - 1.25637984x_{12} + 4.25914509x_{13} - 2.2546124x_{14} - 1.29384865x_{15} \\ & + 1.97365249x_{16} - 0.57078926x_{17} + 2.10366704x_{18} - 4.25139286x_{19} + 10.30095710472393 \end{aligned}$$

## 5-2. 회귀분석 결과

도출된 회귀식을 바탕으로 대학 선호도를 예측하고 대학순위를 산정함

$$\begin{aligned}\hat{y} = & -0.91725493x_1 - 4.67228452x_2 - 0.78279637x_3 + 2.2653429x_4 + 13.50522401x_5 \\ & - 1.82347648x_6 - 4.68199861x_7 - 1.32967242x_8 + 0.04130948x_9 - 3.23905999x_{10} \\ & + 2.29747415x_{11} - 1.25637984x_{12} + 4.25914509x_{13} - 2.2546124x_{14} - 1.29384865x_{15} \\ & + 1.97365249x_{16} - 0.57078926x_{17} + 2.10366704x_{18} - 4.25139286x_{19} + 10.30095710472393\end{aligned}$$



위 분석 방식의 적절성 제고

## 6. 한계

### 1. 상관관계 분석의 유의성

- 회귀분석의 독립변수 선정 근거로 사용한  
상관관계 분석의 유의성 판단 X
- 상관관계 분석의 결과가 단순히 외부 요인이나  
우연의 개입일 가능성 有
- 상관관계의 정도 뿐만 아니라 **상관관계가**  
**통계적으로 유의미한지** 검토할 필요 있음

### 2. 독립변수 간의 관계 배제

- 독립변수와 종속변수의 관계에 영향을 주는  
제3의 요인이 또다른 독립변수로 존재하여  
다중공선성\*을 일으킬 수 있음
- 따라서, 독립변수 간의 관계를 확인하여  
변수 간 **독립성**을 보장해야 함

### 3. 선형회귀분석의 유의성

- 최선의 회귀식을 도출했다 하더라도  
그 회귀식은 유의미하지 않을 수 있다.
- 회귀식의 도출에서 끝나는 것이 아니라  
분석이 유의미했는지 평가할 필요가 있다.

다중공선성\*

회귀분석에 사용된 모형의 일부 독립변수가 다른 독립변수와 상관 정도가 높아,  
데이터 분석 시 부정적인 영향을 미치는 현상



## 7. 한계 극복을 위한 방법 제시

1

 **$p$ -value** 검정

$p$ -value 값 확인하여  
상관관계의 유의성 검증

2

 **$VIF$**  이용

독립변수 간의 독립성 보장,  
다중공선성 발생 방지

3

**회귀분석 결과 확인**

선형회귀분석과  
도출된 회귀식의 유효성 확인

## 7-1. 상관관계 유의성 검사

### 통계적 유의성

- “통계적으로 유의하다.” = 확률적으로 어떤 실험의 결과가 단순한 우연이라고 생각되지 않을 정도로 의미가 있다.
- 보통 가설 검정을 통해 확인함

### 가설 검정

- 귀무가설( $H_0$ ) : 거짓이 명확히 규명될 때까지 참인 것으로 인정되는 가설
- 대립가설( $H_1$ ) : 귀무가설이 거짓이라면 참이 되는 가설, 즉, 귀무가설을 부정하는 새로운 가설
- 검정통계량 : 귀무가설의 타당성을 입증하기 위하여 표본으로부터 얻은 적당한 통계량을 선택한 것
  - ⇒ 선정된 검정통계량을 이용하여 귀무가설( $H_0$ )의 주장이 참이라는 결론을 얻는다면 귀무가설을 채택
  - ⇒ 반면, 대립가설( $H_1$ )의 주장이 참이라는 결론에 도달하면 귀무가설( $H_0$ )을 기각하고 대립가설을 채택

## 7-1. 상관관계 유의성 검사

유의 수준( $\alpha$ ) ; 제 1종 오류를 범할 확률

- 보편적으로 유의수준  $\alpha$  값은 0.01, 0.05, 0.1을 많이 사용

	귀무가설( $H_0$ )이 참	대립가설( $H_1$ )이 참
$H_0$ 채택	올바른 결정	제 2종 오류
$H_1$ 채택	제 1종 오류	올바른 결정

<참조>

$p$ -값 ( $p$ -value)

귀무가설이 사실이라는 전제 아래, 표본에서 관찰된 검정통계량 관찰 값에 의하여 귀무가설을 기각시킬 수 있는 가장 작은 유의수준

⇒  $p$ -값이 유의수준( $\alpha$ )보다 작으면 귀무가설에 대한 신빙성이 떨어지고, 결국 귀무가설을 기각한다.

⇒  $p$ -값이 유의수준( $\alpha$ )보다 크면 귀무가설에 대한 신빙성이 높아지고, 결국 귀무가설을 채택한다.

예) 유의수준( $\alpha$ ) = 0.01,  $p$ -값 = 0.02  $\rightarrow$  유의수준( $\alpha$ ) <  $p$ -값이므로 귀무가설을 채택한다.

## 7-1. 상관관계 유의성 검사

### $p$ -value 이용한 통계적 유의성 검사

파이썬 코드를 통해 각 항목별 상관계수와  $p$ -value를 도출하여 유의성을 검사함

귀무가설( $H_0$ ) : 두 변수는 서로 상관관계가 없다. (상관계수  $r = 0$ )

대립가설( $H_1$ ) : 두 변수는 서로 상관관계가 있다. (상관계수  $r \neq 0$ )

- 유의수준( $\alpha$ )는 디폴트 값인 0.05를 사용
- 따라서 도출된  $p$ -value가 0.05 미만이라면 대립가설을 채택하고 상관계수가 통계적으로 유의미하다고 판단함.
- 반대로 도출된  $p$ -value가 0.05 이상이라면 귀무가설을 채택하고 상관계수가 통계적으로 유의미하지 않다고 판단함.
- 예시

상관계수 : 0.731553123841345

$p$ -value : 0.0005599306029232998

## 7-1. 상관관계 유의성 검사

### 상관관계 유의성 검사 결과

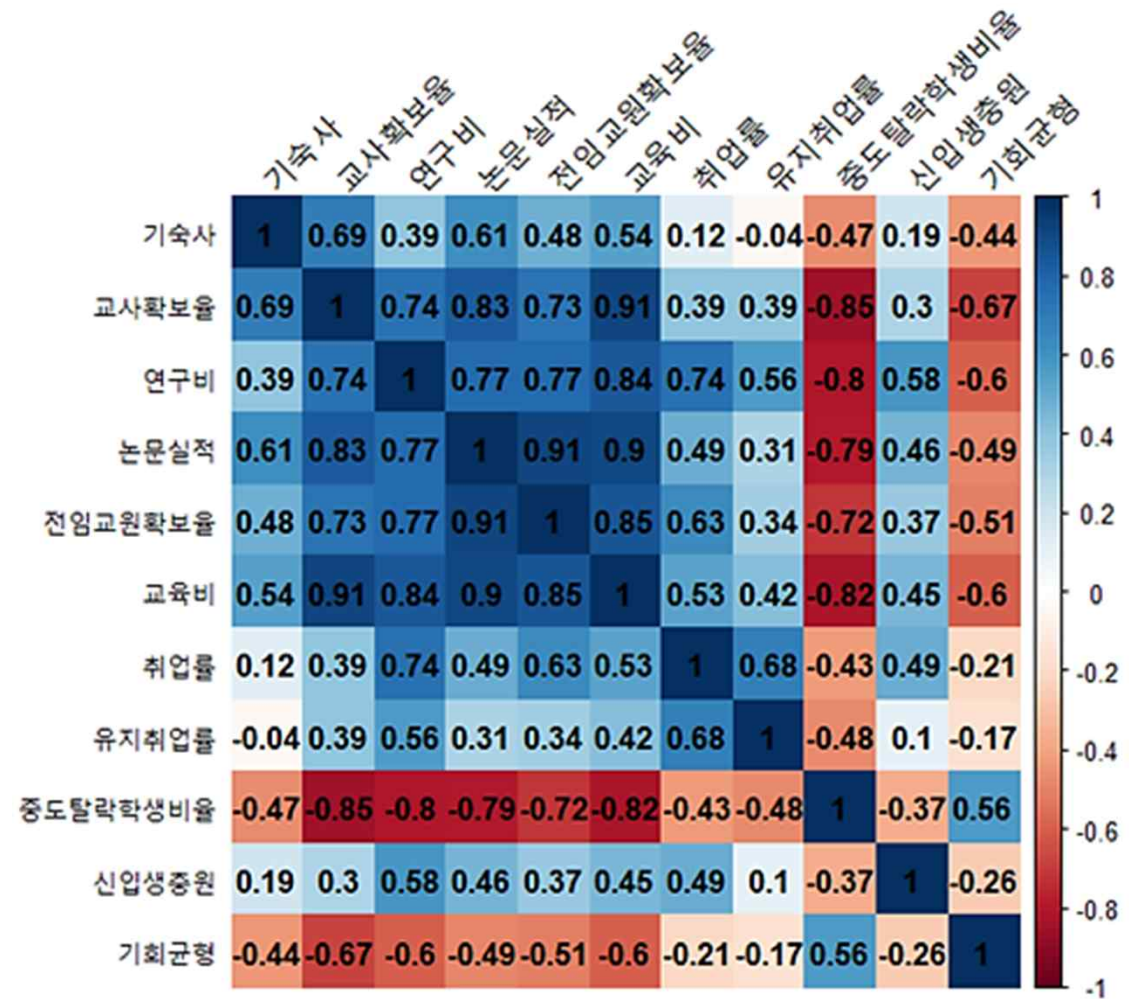
- 고등학생의 대학 선호도와 상관관계가 거의 없거나 통계적으로 유의하지 않은 항목  
 [ ‘교지 확보율’, ‘장학금’, ‘산업체 경력 교원 고용 비율’, ‘외국인교수 보유 현황’, ‘외국인학생 등록 현황’,  
 ‘외국 대학과의 학점 교류 현황’, ‘졸업생 창업 현황’, ‘학생 현장실습 참여 현황’, ‘창업교육 지원 현황’, ‘입학 경쟁률’ ]
- ▶ 총 21개 항목 중 위 10가지 항목을 제거하여 아래 11개 항목을 이후 분석에 활용
- 고등학생의 대학 선호도와 상관관계가 있고 통계적으로 유의성이 증명된 항목  
 [ ‘기숙사 수용률’, ‘교사시설 확보율’, ‘연구비 지원’, 전임교원의 연구 실적’, ‘전임교원 확보율’, ‘교육비 투자’,  
 ‘졸업생 취업률’, ‘졸업생 유지취업률’, ‘중도탈락학생비율’, ‘신입생 충원률’, ‘기회균형 전형 선발 비율’ ]

## 7-2. 변수간 독립성 보장

### 독립변수 간의 상관계수

상관계수가 큰 경우가 존재

- ⇒ 변수간 독립성이 보장되지 않는다는 것을 암시
- ⇒ 회귀분석 진행 시 다중공선성 발생 가능성이 큼



## 7-2. 변수간 독립성 보장

분산팽창요인 (Variation Inflation Factor, VIF) ; 다중공선성을 파악하기 위한 수치적 지표

$$VF_i = \frac{1}{1-R_i^2} \quad (0 < R_i^2 < 1, R_i^2 = i \text{ 번 째 독립변수에 대한 회귀모델의 설명력 (결정계수)})$$

- 독립변수 중 하나를 종속변수로 두고 나머지 독립변수를 독립변수로 두어 회귀분석을 실시한다.
- 결정계수  $R_i^2$  값은 0에서 1의 값을 가지므로  $VF_i$  값은 1에서 무한대의 범위를 갖는다.
- 결정계수  $R_i^2$  값이 0에 가깝다는 것은 독립변수  $x_i$  가 다른 독립변수들과 상관성이 거의 없다는 것을 의미하고,
- 결정계수  $R_i^2$  값이 1에 가깝다는 것은 독립변수  $x_i$  가 다른 독립변수들과 상관성, 즉, 다중공선성이 크다는 것을 의미한다.

따라서,  $VF_i$  값이 1에 가까울수록 다중공선성의 정도가 작은 것이며, 반대로 값이 커질수록 다중공선성의 정도가 큰 것을 의미한다. (절편 제외)

보통  $VF_i$  값이 10 이상인 경우 해당변수가 다중공선성을 유발한다고 판단한다.

## 7-2. 변수간 독립성 보장

Out[48]:

	VIF	features
0	235.7	Intercept
1	3.6	기숙사
2	25.8	교사확보율
3	13.4	연구비
4	17.6	논문실적
5	20.3	전임교원확보율
6	23.9	교육비
7	12.6	취업률
8	4.9	유지취업률
9	9.0	중도탈락학생비율
10	3.5	신입생충원
11	3.6	기회균형

▲ VIF Factor를 활용하여 분석한 결과

### VIF 계산과 회귀분석에 사용할 변수 선택

- 파이썬 코드를 통해 11개 항목에 대해 VIF 계산
- 변수들을 제거/대체해가며 최적의 독립변수 조합을 선택

Out[49]:

	VIF Factor	features
0	33.291990	Intercept
1	2.743179	연구비
2	2.743179	중도탈락학생비율

∴ 회귀분석 모델의 성능 및 다중공선성 발생 여부를 고려하여

(1) '대학의 전임교원 연구비 지원', (2) '대학의 중도탈락 학생 비율'

위 두 가지를 독립변수로 최종 결정



## 7-3. 선형회귀분석과 도출된 회귀식의 유효성 확인

선형회귀분석과 도출된 회귀식의 유효성을 나타내는 지표

### 1. 결정계수 (R-Squared)

- 회귀분석의 정확도와 설명력을 알려주는 지표

$$R^2 = \frac{SSR}{SST} = \frac{\sum_i^n (\hat{y}_i - \bar{y})^2}{\sum_i^n (y_i - \hat{y}_i)^2 + \sum_i^n (\hat{y}_i - \bar{y})^2}$$

- 종속변수의 총 변량 중 회귀식으로 설명 가능한 변량의 비율
- 결정계수( $R^2$ )는 0에서 1 사이의 값을 갖는다.
- 결정계수( $R^2$ )가  
1에 가까울수록 회귀식의 정확도와 설명력 ↑  
0에 가까울수록 회귀식의 정확도와 설명력 ↓

### 2. 회귀식의 절편에 대한 $p$ -value

- 회귀식의 절편에 대한 통계적 유의성을 검증
- 귀무가설( $H_0$ ) : 회귀식의 상수항이 유의미하지 않다.  
(절편  $\beta_0 = 0$ )
- 대립가설( $H_1$ ) : 회귀식의 상수항이 유의미하다.  
(절편  $\beta_0 \neq 0$ )
- 유의수준 0.05에서,  
(1) 절편의  $p$ -value < 0.05 → 대립가설 채택  
(2) 절편의  $p$ -value ≥ 0.05 → 귀무가설 채택

### 3. 기울기에 대한 $p$ -value

- 각 독립변수에 대한 기울기의 통계적 유의성을 검증
- 귀무가설( $H_0$ ) : 해당 독립변수는 종속변수를 설명할 수 없다.  
(기울기  $\beta_i = 0$ )
- 대립가설( $H_1$ ) : 해당 독립변수는 종속변수를 설명할 수 있다.  
(기울기  $\beta_i \neq 0$ )
- 유의수준 0.05에서,  
(1) 기울기의  $p$ -value < 0.05 → 대립가설 채택  
(2) 기울기의  $p$ -value ≥ 0.05 → 귀무가설 채택

## 7-3. 선형회귀분석과 도출된 회귀식의 유효성 확인

### 선형회귀분석 결과

#### 1. 결정계수( $R^2$ ) = 0.907

‘대학의 중도탈락 학생 비율’ 항목과 ‘대학의 연구비 지원’ 항목은  
고등학생의 대학 선호도의 약 91%를 설명함

#### 2. 회귀식의 절편에 대한 $p$ -value = 0.000

회귀식의 상수항은 통계적으로 유의미함

#### 3. 기울기에 대한 $p$ -value = 0.004 & 0.001

각 독립변수의 기울기는 통계적으로 유의미함

‘대학의 중도탈락 학생 비율’ 항목과 ‘대학의 연구비 지원’ 항목은  
고등학생의 대학 선호도를 설명할 수 있음

### OLS Regression Results

Dep. Variable:	rank	R-squared:	0.907		
Model:	OLS	Adj. R-squared:	0.894		
Method:	Least Squares	F-statistic:	72.97		
Date:	Sat, 17 Oct 2020	Prob (F-statistic):	1.87e-08		
Time:	23:29:06	Log-Likelihood:	-27.026		
No. Observations:	18	AIC:	60.05		
Df Residuals:	15	BIC:	62.72		
Df Model:	2				
Covariance Type:	nonrobust				
	coef	std err	t P> t  [0.025 0.975]		
Intercept	11.9546	1.618	7.389 0.000	8.506 15.403	
중도탈락학생비율	-5.8958	1.732	-3.403 0.004	-9.588 -2.203	
연구비	6.4231	1.499	4.285 0.001	3.228 9.618	
Omnibus:	1.128	Durbin-Watson:	2.095		
Prob(Omnibus):	0.569	Jarque-Bera (JB):	0.777		
Skew:	-0.018	Prob(JB):	0.678		
Kurtosis:	1.983	Cond. No.	12.0		

### 결과 해석

해당 선형회귀분석은 통계적으로 유의미하며, 도출된 회귀식은 ‘고등학생의 대학 선호도’에 대해 약 91%의 높은 설명력을 가진다.

## 7-3. 선형회귀분석과 도출된 회귀식의 유효성 확인

### 도출된 회귀식

$$\hat{y} = 6.4231x_1 - 5.8958x_2 + 11.9546$$

( $\hat{y}$  = 선호도 응답률 예측값,  $x_1$  = 대학의 중도탈락 학생 비율,  $x_2$  = 대학의 전임교원 1인당 평균 연구비 지원)

### 실제 응답률과 예측한 응답률 비교

Out [57]:

	대학명	실제값	예측값
0	성균관대학교	15.1	15.755092
1	이화여자대학교	11.5	11.889981

## 7-3. 선형회귀분석과 도출된 회귀식의 유효성 확인

독립변수가 2개인 회귀선이 최선의 회귀선인 이유

### 1 설명력의 부족 (기숙사, 취업률, 유지취업률, 기회균형, 신입생충원 총 5개 항목)

- 위의 항목들은 ‘고등학생의 대학 선호도 응답률’과의 상관계수가 다른 항목들보다 비교적 낮아 선호도를 설명하기에 부족함.

↳ 각각 [ 0.51, 0.59, 0.55, -0.61, 0.55 ] VS [ 0.83, 0.91, 0.84, 0.73, 0.83, -0.89 ]

- 각 항목의 ‘고등학생의 대학 선호도 응답률’에 대한 설명력을 확인하기 위해 단순선형회귀분석을 실시한 결과

회귀모델의 결정계수( $R^2$ )가 다른 항목들보다 현저히 낮음

↳ 각각 [ 0.211, 0.311, 0.259, 0.328, 0.257 ] VS [ 0.672, 0.825, 0.693, 0.506, 0.657, 0.780 ]

따라서, 위 5개 항목을 독립변수로 사용하면 오히려 종속변수에 대한 예측의 정확도가 떨어진다.

## 7-3. 선형회귀분석과 도출된 회귀식의 유효성 확인

독립변수가 2개인 회귀선이 최선의 회귀선인 이유

### 2 유사한 정보를 내포 (‘대학의 전임교원 1인당 평균 논문실적’ 항목)

- 교수진의 역량이라는 점에서 ‘대학의 전임교원 1인당 평균 연구비 지원’ 항목과 유사한 정보를 내포함
- 두 항목의 상관계수는 0.77로 강한 양의 상관관계를 보임
- 두 항목을 동시에 선형회귀분석의 독립변수로 사용할 때와 ‘연구비’ 항목만을 선형회귀분석의 독립변수로 사용할 때의 결정계수( $R^2$ ) 상승폭이 작음

(1) ‘연구비’ 항목만을 독립변수로 사용하는 경우

결정계수( $R^2$ ) : 0.825

(2) ‘연구비’ 항목과 ‘논문실적’ 항목을 동시에 독립변수로 사용하는 경우

결정계수( $R^2$ ) : 0.860, 논문실적의 기울기  $p$ -value : 0.03

‘전임교원 1인당 평균 연구비 지원’ 항목과 ‘전임교원 1인당 평균 논문실적’ 항목은 유사한 정보를 내포하므로  
둘 중 하나의 항목만을 추가해도 종속변수에 대한 충분한 설명이 가능하다.

## 7-3. 선형회귀분석과 도출된 회귀식의 유효성 확인

독립변수가 2개인 회귀선이 최선의 회귀선인 이유

### 3 변수간 강한 연관성 (‘교사(校舍)시설 확보율’, ‘전임교원확보율’, ‘학생 1인당 평균 교육비 투자’ 총 3개 항목)

- 3개 항목은 학생의 전반적인 만족도라는 점에서 ‘**대학의 중도탈락 학생 비율\***’ 항목과 **연관성이 강함**

\*대학의 중도탈락 학생 비율 : 대학에 소속되어 있는 재적학생 중 자퇴, 미복학, 미등록 등의 이유로 학업을 중단한 학생의 비율

- 논문 자료를 통해 ‘대학의 중도탈락 학생 비율’과 위의 세 가지 항목의 관계를 확인할 수 있음

<참조>

“교육 여건의 또 다른 측면에서 전임교원 1인당 학생수가 많을수록 중도탈락율에 부정적인 영향을 미치는 것으로 나타났다. (중략)

대학생의 중도탈락을 사전에 예방하고 학생들을 실질적으로 지원할 수 있도록 교육 여건이 개선될 필요가 있다.”

“대학 재정 및 교육비 측면에서 살펴보면 학생 1인당 교육비가 높은 것은 학생의 중도탈락율을 낮추는 데 긍정적인 영향을 미치는 것으로 나타났다. 학생 1인당 교육비는 교육에 투자되는 총교육비를 학생수로 나눈 것으로, 교육의 질과 밀접하게 관련되어 있다고 볼 수 있다.”

(출처 : 「대학생의 중도탈락에 영향을 미치는 대학수준 요인 분석」, 정제영 외 2명, 아시아교육연구 16권 4호, 2015)

## 7-3. 선형회귀분석과 도출된 회귀식의 유효성 확인

독립변수가 2개인 회귀선이 최선의 회귀선인 이유

### 3 변수간 강한 연관성 (‘교사(校舍)시설 확보율’, ‘전임교원 확보율’, ‘학생 1인당 평균 교육비 투자’ 총 3개 항목)

- ‘교사(校舍)시설\* 확보율’, ‘전임교원 확보율’ 항목은 학생의 교육 여건을 나타낸다는 점에서 ‘중도탈락 학생 비율’ 항목과 상관성이 크다.  
↳ 교육을 위한 학교의 각종 시설 및 건물
- ‘학생 1인당 평균 교육비 투자’ 항목은 교육의 질을 나타낸다는 점에서 ‘중도탈락 학생 비율’ 항목과 상관성이 크다.
- 상관관계 분석을 통해 위 세 항목이 ‘중도탈락 학생 비율’ 항목과 높은 상관성을 갖는다는 것을 알 수 있음  
↳ 각각 [ -0.85, -0.82, -0.72 ]
- 위 세 가지 항목을 ‘대학의 중도탈락 학생 비율’ 항목과 함께 선형회귀분석의 독립변수로 활용하는 것은  
‘대학의 중도탈락 학생 비율’ 항목 하나만을 독립변수로 활용할 때보다 회귀모델의 결정계수( $R^2$ )를 감소시킴.  
→ 다중공선성 발생

‘대학의 교사(校舍)시설 확보율’, ‘대학의 전임교원 확보율’, ‘대학의 학생 1인당 평균 교육비 투자’ 항목은 ‘대학의 중도탈락 학생 비율’ 항목과 강한 연관성을 보이므로 이 항목들을 동시에 회귀분석의 독립변수로 선택할 수 없다.

## 7-3. 선형회귀분석과 도출된 회귀식의 유효성 확인

‘대학의 전임교원 1인당 평균 연구비 지원’, ‘대학의 중도탈락 학생 비율’을 회귀분석의 독립변수로 결정

- (1) 위의 두 항목을 동시에 사용할 때 회귀모델의 결정계수( $R^2$ )가 가장 높다.
- (2) 다중공선성이 발생하지 않고, 회귀분석 시 각 회귀계수가 통계적 유의성을 보인다.



## 8. 결론

도출된 회귀식을 통해 대학 선호도를 예측하고 새로운 대학 순위를 도출

(1) 도출된 회귀식

$$\hat{y} = 6.4231x_1 - 5.8958x_2 + 11.9546$$

(2) 2020년 서울 소재 종합 대학의 데이터를 도출한 회귀식에 대입하여 최종 순위를 산출

## 8. 결론

순위	대학명	선호도 예측값	순위	대학명	선호도 예측값
1	서울대학교	18.144867	16	숙명여자대학교	10.802008
2	고려대학교	17.056695	17	광운대학교	10.725250
3	성균관대학교	16.973942	18	상명대학교	10.455269
4	연세대학교	16.870038	19	명지대학교	9.733432
5	한양대학교	14.670885	20	홍익대학교	9.610590
6	이화여자대학교	13.452743	21	성신여자대학교	9.279212
7	중앙대학교	13.388297	22	한국외국어대학교	9.253159
8	건국대학교	12.673369	23	세종대학교	9.247680
9	서울시립대학교	12.534834	24	덕성여자대학교	9.126780
10	서강대학교	12.366981	25	서울여자대학교	8.961754
11	경희대학교	12.225012	26	삼육대학교	7.995370
12	동국대학교	11.759016	27	동덕여자대학교	7.743359
13	국민대학교	11.611041	28	한성대학교	7.298289
14	송실대학교	11.406155	29	서경대학교	6.593473
15	서울과학기술대학교	11.377959			

## 8. 결론

### 결론

- (1) 대학 순위를 개발하기 위해 **고등학생의 대학 선호도에 영향을 미치는 대학의 특성과 그 정도를 파악하고자 함**
- (2) 수집한 대학 데이터의 각 항목과 고등학생의 대학 선호도 응답률 간의 **상관관계를 분석하여 대학 선호도에 영향을 주는 요인을 도출함**
- (3) 상관관계 분석 이후, **다중선형회귀분석을 사용하여 대학 선호도에 영향을 주는 특성들을 최종 결정하고 각 값이 대학 선호도에 얼마만큼의 영향을 주는지 파악함**
- (4) 다중선형회귀분석을 통해 도출된 회귀식을 사용하여 **2020년 서울 소재 종합 대학의 선호도를 예측하고 이를 바탕으로 대학 진학 예정자를 위한 새로운 대학 순위를 제시함**
- (5) **대학 평가기준** : ‘대학의 전임교원 1인당 평균 연구비 지원’, ‘대학의 중도탈락 학생 비율’
- (6) **평가기준에 대한 배점** : 회귀모델을 사용하여 배점 부여

## 8. 결론

### 한계

- (1) 수집 가능한 자료들의 한계로 평가 기준이 지엽적일 수밖에 없었음
- (2) 위 과정을 통해 선호도에 영향을 주는 요인을 찾았으나 인과성의 증명에는 더 주의 깊은 연구가 필요함

## 출처

### \* 데이터 수집

‘대학알리미’ 대학 공시 데이터

<https://www.academyinfo.go.kr/index.do>

2014 내일신문 고등학생 대학 선호도조사

<http://www.newscj.com/news/articleView.html?idxno=267617>

### \* 참고자료

《평학인증을 위한 확률과 통계》, 이재원 외 1명, 북스힐, 2019

《파이썬 데이터 사이언스 핸드북》, 제이크 밴더플래스, 김정인 옮김, 위키북스, 2017

《대학생의 중도탈락에 영향을 미치는 대학수준 요인 분석》, 정제영 외 2명, 아시아교육연구 16권 4호, 2015



**THANK YOU!**