

KT AIVLE School

1차 미니프로젝트_3반_10조

AI 03반 10조

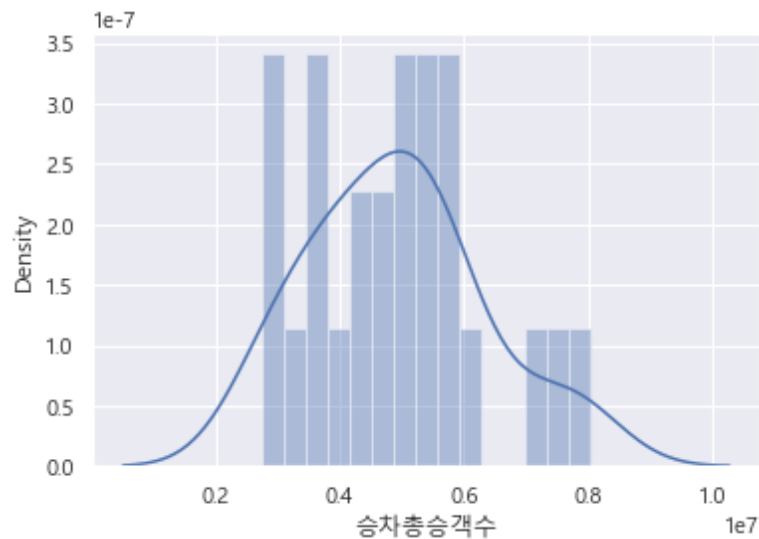
가설 수립

- ✓ 가설 1 : 승차 총 승객 수가 많은 구를 지나가는 노선이 많을 것이다.
- ✓ 가설 2 : 평균 이동 시간(평균)이 길수록 노선 수가 적다.
- ✓ 가설 3 : (승차총승객수 / 정류장수)와 평균 이동 시간_mean은 양의 상관관계를 가질 것이다.

단변량 분석

✓ 가설 1 : 승차 총 승객 수가 많은 구를 지나가는 노선이 많을 것이다.

```
var = '승차총승객수'
target = '노선수'
sns.distplot(df[var], bins=15)
plt.show()
```



```
df[var].isna().sum()
```

0

```
df[var].describe()
```

```
count    2.500000e+01
mean     4.912037e+06
std      1.421263e+06
min      2.749448e+06
25%      3.792541e+06
50%      4.942197e+06
75%      5.641742e+06
max      8.030483e+06
Name: 승차총승객수, dtype: float64
```

이변량 분석

✓ 가설 1 : 승차 총 승객 수가 많은 구를 지나가는 노선이 많을 것이다.

▪ X 와 Y의 관계

X : 승차 총 승객 수

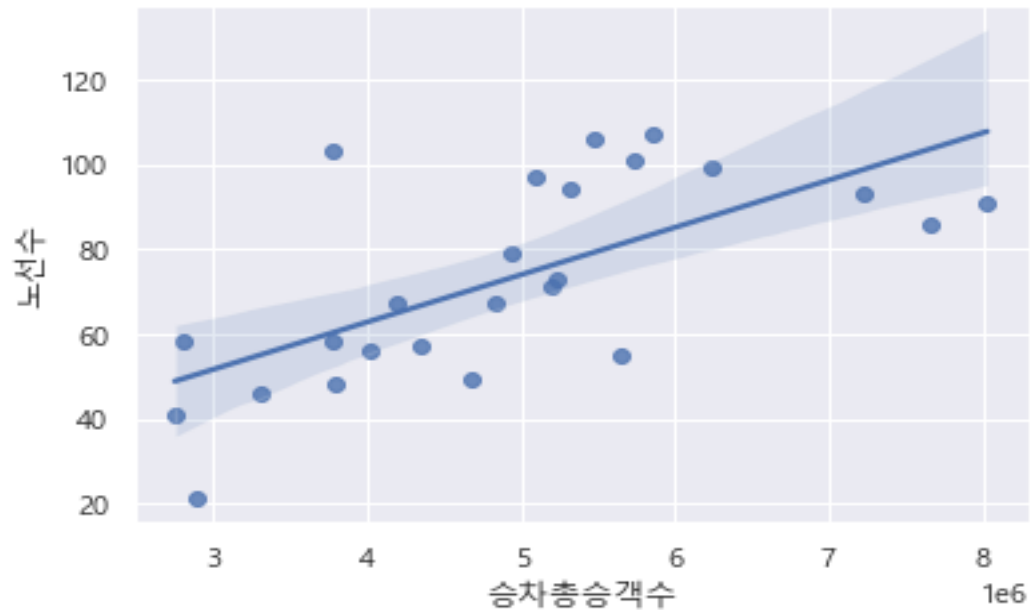
Y : 노선 수

상관계수 : 0.665

P-value : 0.00028

```
spst.pearsonr(df[var], df[target])
```

(0.6654877832095702, 0.0002830862285771981)



단변량 분석

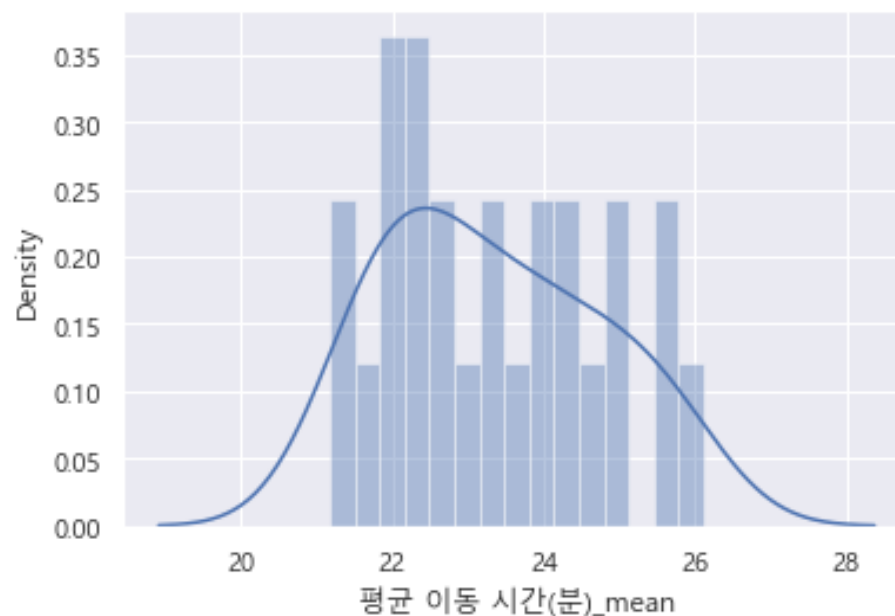
✓ 가설 2 : 평균 이동 시간(평균)이 길수록 노선 수가 적다.

```
df['평균 이동 시간(분)_mean'].describe()
```

```
count    25.000000
mean     23.333424
std       1.433409
min      21.166269
25%      22.215130
50%      23.233772
75%      24.394485
max      26.108477
Name: 평균 이동 시간(분)_mean, dtype: float64
```

```
df['노선수'].describe()
```

```
count    25.000000
mean     72.920000
std      23.808122
min      21.000000
25%      56.000000
50%      71.000000
75%      94.000000
max     107.000000
Name: 노선수, dtype: float64
```



이변량 분석

✓ 가설 2 : 평균 이동 시간(평균)이 길수록 노선 수가 적다.

- X와 Y의 관계를 살펴봅시다.

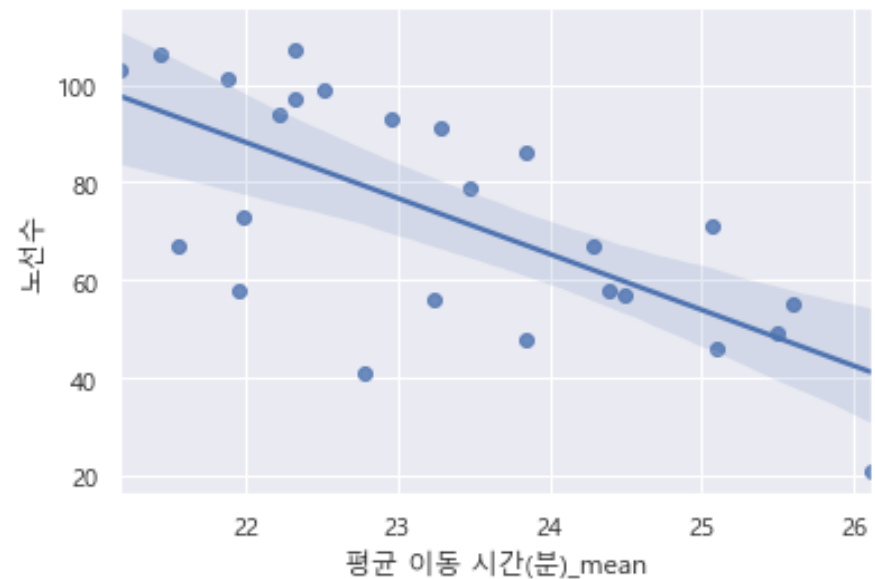
X : 평균 이동 시간 _ 평균

Y : 노선 수

상관계수 : -0.6874

P-value : 0.00015

```
sns.regplot(x=df['평균 이동 시간(분)_mean'], y=df[target])
plt.show()
spst.pearsonr(df['평균 이동 시간(분)_mean'], df[target])
```



: (-0.6874048303585873, 0.00014680687151646322)

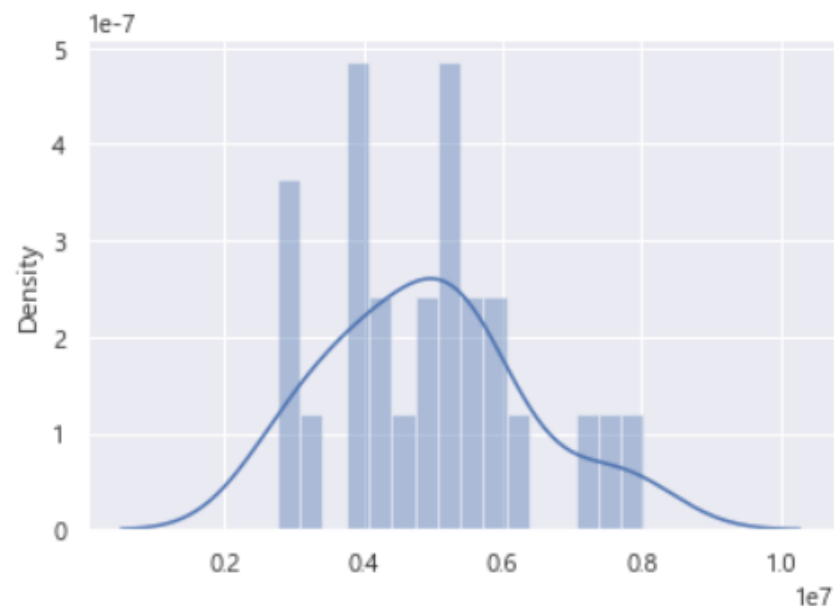
단변량 분석

✓ 가설 3 : (승차총승객수 / 정류장수)와 평균 이동 시간_mean은 양의 상관관계를 가질 것이다.

```
df['정류장당 승차승객수'] = df['승차총승객수']/df['정류장수']
df[['자치구', '정류장당 승차승객수']]
```

	자치구	정류장당 승차승객수		자치구	정류장당 승차승객수
0	종로구	15670.300860	12	서대문구	12858.550439
1	중구	21217.275281	13	마포구	9054.201068
2	용산구	12972.065015	14	양천구	12567.937304
3	성동구	6453.517241	15	강서구	8270.464664
4	광진구	10220.996283	16	구로구	10169.129630
5	동대문구	17126.029412	17	금천구	10946.834783
6	종랑구	9980.371053	18	영등포구	12343.817204
7	성북구	10359.508306	19	동작구	12133.045662
8	강북구	11683.246973	20	관악구	16428.796137
9	도봉구	9204.192201	21	서초구	12035.550000
10	노원구	8469.445525	22	강남구	16093.152305
11	은평구	10382.554890	23	송파구	13594.559036
			24	강동구	7832.121951

```
var = '승차총승객수'
target = '정류장수'
sns.distplot(df[[var]], bins=16)
plt.show()
```



이변량 분석

✓ 가설 3 : (승차총승객수 / 정류장수)와 평균 이동 시간_mean은 양의 상관관계를 가질 것이다.

- X와 Y의 관계

X : 승차총승객수 / 정류장수

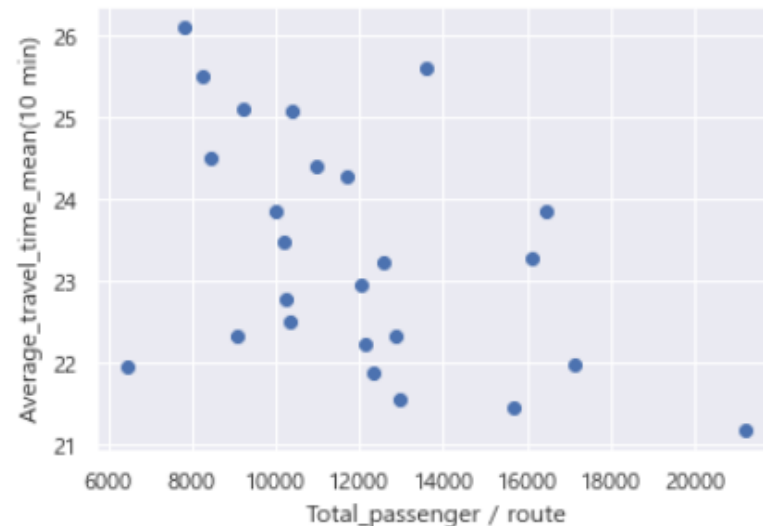
Y : 평균 이동 시간의 평균

상관계수 : -0.4535

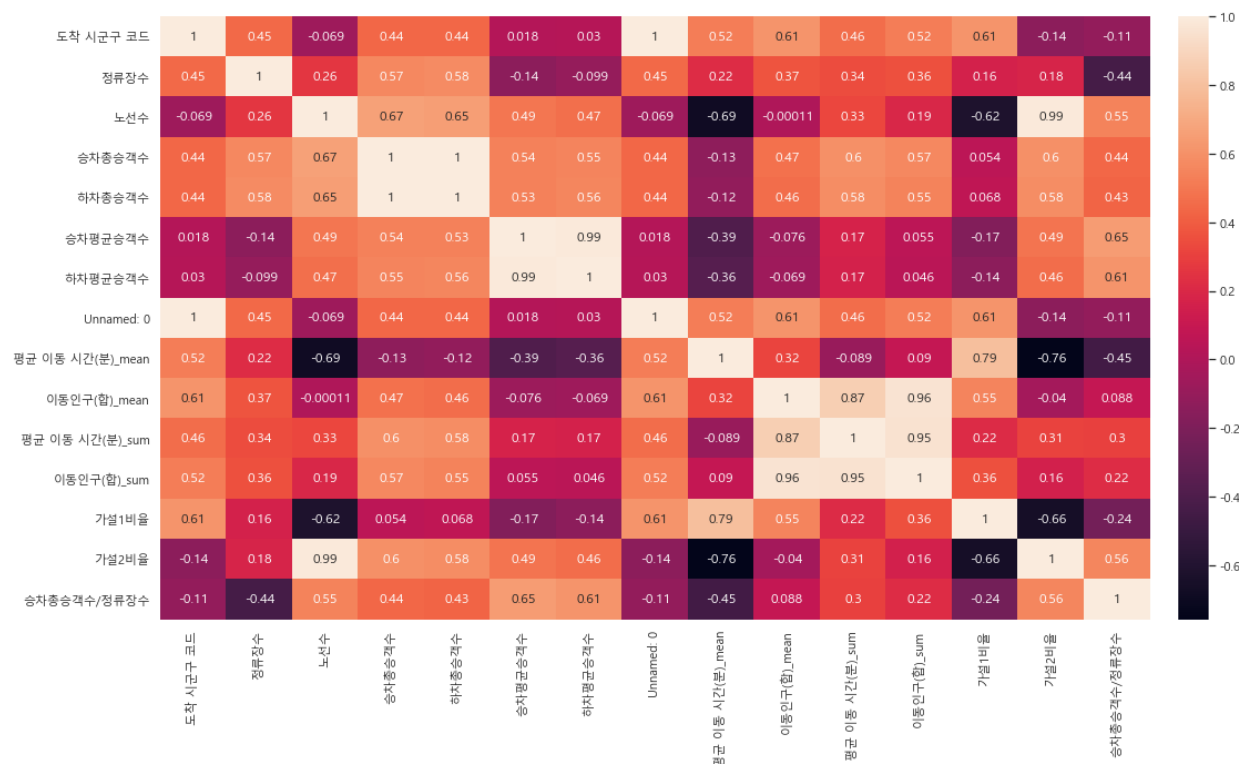
P-value : 0.02279

```
a2 = df['승차총승객수']/df['정류장수']
plt.scatter(a2, df['평균 이동 시간(분)_mean'])
plt.xlabel('Total_passenger / route')
plt.ylabel('Average_travel_time_mean(10 min)')
a21 = spst.pearsonr(a2, df['평균 이동 시간(분)_mean'])
a21
```

(-0.45352924781086634, 0.02278553520964107)



관계 정리



1) 강한 관계

: (승차총승객수 / 정류장수) ~

평균 이동 시간_mean

: 승차 총 승객수 ~ 노선수

: 평균 이동 시간 ~ 노선수

2) 약한 관계

: 노선수 ~ 이동인구(합)_sum

: 정류장수 ~ 하차평균승객수

가설 검증 과정



미니 프로젝트 #1

- Mural :

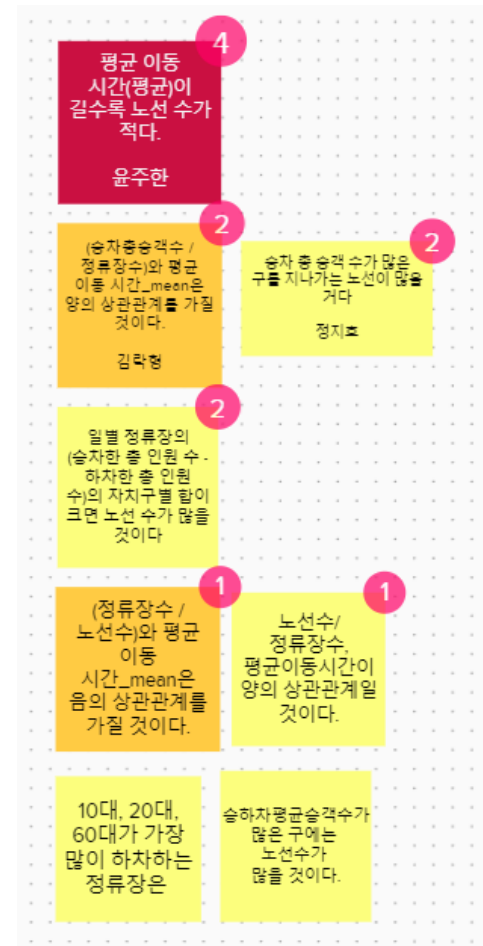
MURAL

<https://app.mural.co/invitation/mural/htaivle29596/1660702495841?sender=u23b114c16816bfe315db6854&key=1f38fbb5-7c...>

- 1 문제 해결 방법론
- 2 조별 토론
- 3 가설 공유

- 최종 제출 파일

2.1 조별 제출물_3반_10조.ipynb 452.3KB



가설 검증 과정

가설 1) 승차 총 승객 수가 많은 구를 지나가는 노선이 많을 것이다.

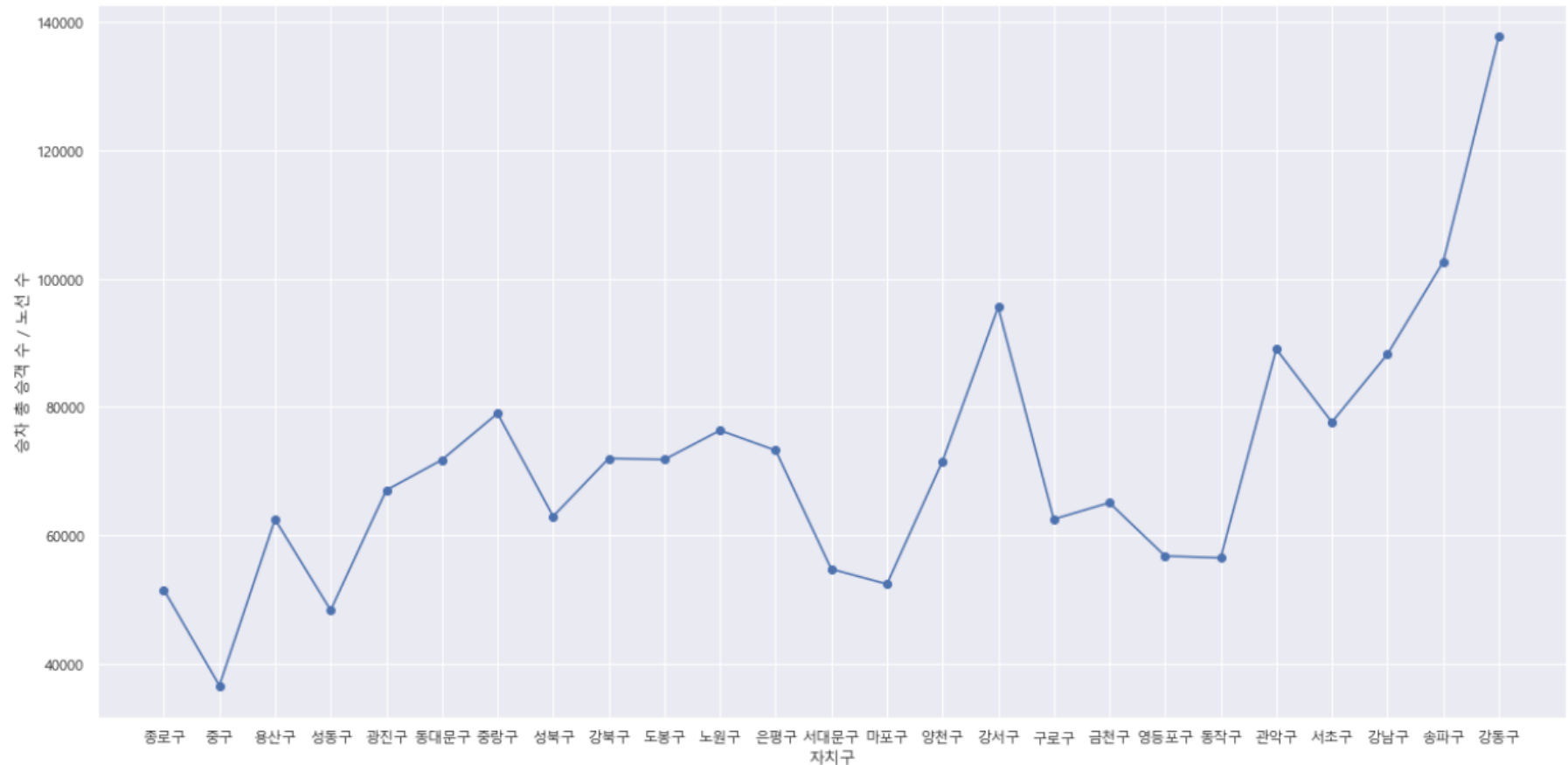
- 승차 총 승객수가 많은 자치구는 일반적으로 출발지는 동일하나 도착지가 다양할 것입니다. 버스의 수요가 많은 자치구는 버스가 많이 필요할 것입니다.
- 위의 이유로 승차 총 승객수가 많은 구는 노선이 많을 것이라고 판단하여 가설을 수립했습니다.
- 이변량 분석 결과에서 나타났듯이 강한 상관관계를 보였습니다.
- 승차 총 승객수가 많지만 노선이 적은 자치구에 노선의 보충이 이루어져야 할 것이라고 결론지을 수 있습니다.

자치구		가설1비율			
0	강동구	137621.571429	13	광진구	67059.707317
1	송파구	102577.127273	14	금천구	65114.793103
2	강서구	95532.306122	15	성북구	62994.181818
3	관악구	89021.151163	16	구로구	62559.455696
4	강남구	88247.065934	17	용산구	62536.970149
5	종로구	79011.270833	18	영등포구	56830.445545
6	서초구	77648.709677	19	동작구	56534.829787
7	노원구	76373.596491	20	서대문구	54799.056075
8	은평구	73262.816901	21	마포구	52458.360825
9	강북구	72017.626866	22	종로구	51593.726415
10	도봉구	71832.717391	23	성동구	48401.379310
11	동대문구	71788.561644	24	중구	36666.747573
12	양천구	71592.357143			

정지호 에이블러님

가설 검증 과정

✓ 가설 1 : 승차 총 승객 수가 많은 구를 지나가는 노선이 많을 것이다.



가설 검증 과정

가설 2) 평균 이동 시간(평균)이 길수록 노선 수가 적을 것이다.

- 일반적으로 출발지에서 도착지까지 이동할 수 있는 경우의 수가 많을 때 평균 이동 시간이 짧을 것이라고 판단했습니다.
- 위의 이유를 명제로 봤을 때 Y를 노선수로 설정하기 위해 명제의 대우를 가져왔습니다.
- 평균 이동시간(합)이 아닌 평균 이동시간(평균)을 X로 지정한 이유는 다음과 같습니다.

평균 이동시간(합)의 일부 데이터가 아주 멀거나 아주 짧은 거리를 이동한 데이터일 때 노선수와 관계를 대표하지 못할 것으로 생각했기 때문입니다.

- 위의 결과들을 통해 분석한 결과 평균 이동시간이 긴 지역에서 노선 수가 적기 때문에 (노선 수 / 평균 이동시간)의 비율이 낮은 지역에 노선을 확대하는 것을 고려해 볼 수 있습니다.

자치구		가설2비율			
0	종로구	4.943262			
1	중구	4.866233	13	마포구	4.345426
2	용산구	3.108526	14	양천구	2.410284
3	성동구	2.642278	15	강서구	1.921796
4	광진구	1.799897	16	구로구	3.364507
5	동대문구	3.320304	17	금천구	2.377587
6	종랑구	2.012657	18	영등포구	4.616529
7	성북구	4.398288	19	동작구	4.231350
8	강북구	2.759195	20	관악구	3.607009
9	도봉구	1.832459	21	서초구	4.050638
10	노원구	2.326764	22	강남구	3.908167
11	은평구	2.832506	23	송파구	2.147967
12	서대문구	4.792127	24	강동구	0.804336

윤주한 에이블러님

가설 검증 과정

가설 3) (승차총승객수 / 정류장수)와 평균 이동 시간_mean은 양의 상관관계를 가질 것이다.

- 정류장 당 승차 승객수를 구하기 위해 승차 총 승객수를 정류장 수로 나누었습니다.
- 이를 토대로 평균이동시간의 평균과의 가설을 수립했습니다.
- 예상 가설대로 서로 관계가 있음이 드러났으나 회귀 직선의 기울기가 예상과 달리 음의 상관관계를 보입니다.
- 위 결과를 통해 분석한 결과 단순히 정류장 수만 많은 것은 오히려 정차구간을 늘려 평균 이동 시간을 증가시킴을 알 수 있습니다.

	자치구	승차총승객수/정류장수	평균 이동 시간(분)_mean
3	성동구	6453.517241	21.950757
24	강동구	7832.121951	26.108477
15	강서구	8270.464664	25.496986
10	노원구	8469.445525	24.497542
13	마포구	9054.201068	22.322323
9	도봉구	9204.192201	25.102882
6	중랑구	9980.371053	23.849072
16	구로구	10169.129630	23.480408
4	광진구	10220.996283	22.779081
7	성북구	10359.508306	22.508757
11	은평구	10382.554890	25.066142
17	금천구	10946.834783	24.394485
8	강북구	11683.246973	24.282441

21	서초구	12035.550000	22.959344
19	동작구	12133.045662	22.215130
18	영등포구	12343.817204	21.877909
14	양천구	12567.937304	23.233772
12	서대문구	12858.550439	22.328288
2	용산구	12972.065015	21.553622
23	송파구	13594.559036	25.605607
0	종로구	15670.300860	21.443330
22	강남구	16093.152305	23.284573
20	관악구	16428.796137	23.842468
5	동대문구	17126.029412	21.985940
1	종구	21217.275281	21.166269
23.33342421099137			

김락형 에이블러님

결론

- 어느 구에 버스 시설의 추가가 가장 필요한가요??

(가설1) 승차 총 승객 수가 많은 구를 지나가는 노선이 많을 것이다.

- 강동구는 비슷한 승차 총 승객수를 가진 자치구에 비해 노선 수가 적어 보충이 필요해 보입니다.
- 송파구 또한 승차 총 승객수에 비해 노선이 적어 보충이 필요합니다.

(가설2) 평균 이동 시간(평균)이 길수록 노선 수가 적을 것이다.

- 평균 이동시간의 평균에 비해 강동구(21개의 노선수)에 노선 수가 너무 적습니다.

(가설3) (승차총승객수 / 정류장수)와 평균 이동 시간_mean은 양의 상관관계를 가질 것이다.

- 평균 이동시간과 정류장수를 고려하여 평균 이하의 평균이동시간 (평균이동시간의 평균 = 23.3) 을 가져 정류장을 늘리는 것에 큰 부담이 없고 정류장당 승차 승객수가 가장 많은 중구에 정류장 증설이 필요합니다.

→ 3개 가설 증명에서 중복적으로 나타나는 "**강동구**" 에 노선 증설이 가장 필요합니다.

kt

 AIVLE

 AIVLE
Let's make it possible