

정규성 검정

정규분포 확인방법 : Shapiro 값의 유의확률이나 Kolmogorov-Smirnov 유의확률

각각의 변수가 **0.05 보다 낮은** 유의확률을 보일 때 귀무가설(정규분포이다)
기각하여 정규분포가 충족되지 않는다고 볼 수 있습니다

꼭 정규성검정을 하지 않고 비모수검정을 진행하는 경우도 많이 있습니다.

n 수가 20 명 미만일 때입니다.

10 명 이하일 때는 무조건 비모수검정. 20 명 미만일 때는 정규성검정 이후 비모수검정. 30 명 이상일 때는 일반 모수통계를 진행해도 된다고 합니다.

하지만 30 명 이상이라도 정규분포가 충족되지 않는 경우가 매우 많기 때문에 정규성검정
결과에 따라 비모수검정 or 모수통계 방법을 결정해서 진행

출처 <https://blog.naver.com/idsde/222460677471>

정규성 테스트의 해석

각 테스트는 최소한 두 가지를 반환합니다.

- **통계** : 테스트 통계 분포의 임계 값과 비교하여 테스트 컨텍스트에서 해석 할 수있는 테스트에서 계산 된 수량입니다.
- **p- 값** : 검정을 해석하는 데 사용됩니다.이 경우 표본이 가우스 분포에서 추출되었는지 여부입니다.

이 테스트에서는 표본이 가우스 분포에서 추출되었다고 가정합니다. 기술적으로 이것을 귀무 가설 또는 H_0 이라고합니다. p- 값을 해석하는 데 사용되는 알파 (일반적으로 5 % (또는 0.05))라는 임계 수준이 선택됩니다.

이러한 테스트의 SciPy 구현에서 p- 값을 다음과 같이 해석 할 수 있습니다.

- **$p \leq \alpha$** : 정상이 아닌 H_0 을 거부합니다.
- **$p > \alpha$** : H_0 거부 실패, 정상.

shapiro , normaltest

shapiro 가 노말 분포 검정할 때 가장 엄격하게 한다고 함.

```
from scipy.stats import shapiro , normaltest , anderson , kstest
```

```
normal = []
notnormal = []
for var in num_var :
    stat, p = shapiro(train[var].dropna().values)
    stat, p = normaltest(train[var].dropna().values)
    alpha = 0.05
    if p > alpha :
        normal.append(var)
    else :
        notnormal.append(var)
```

kstest

분포 2 개를 비교하는 것이므로, 어떤 분포에서도 사용 가능하므로, 노말에서도 사용 가능함.

```
normal = []
notnormal = []
for var in num_var :
    stat,p = kstest(train[var].dropna().values, "norm")
    alpha = 0.05
    if p > alpha :
        normal.append(var)
    else :
        notnormal.append(var)
```

anderson

kstest와 같이 어떤 분포라도 비교 가능함. 기존 분포를 알고 있는 경우 ks test 보다 더 정교하게 비교가 가능함.

비교 가능한 분포로는 {'norm','expon','logistic','gumbel','gumbel_l', 'gumbel_r', 'extreme1'}

```
normal = []
notnormal = []
for var in num_var :
    result = anderson(train[var].dropna().values)
    normality = 0
    for i in range(len(result.critical_values)):
        sl, cv = result.significance_level[i], result.critical_values[i]
        if result.statistic < result.critical_values[i]:
            normality +=1
        else :
            normality +=0
    if normality > 2.5 :
        normal.append(var)
    else :
        notnormal.append(var)
```

좀 더 자세한 것은 아래 글 참고!

<https://towardsdatascience.com/6-ways-to-test-for-a-normal-distribution-which-one-to-use-9dcf47d8fa93>



6 ways to test for a Normal Distribution—which one to use?

Find out which approach is the most powerful when testing for a normal distribution!

towardsdatascience.com

<https://machinelearningmastery.com/a-gentle-introduction-to-normality-tests-in-python/>



A Gentle Introduction to Normality Tests in Python

An important decision point when working with a sample of data is whether to use parametric or nonparametric statistical methods. Parametric statistical methods assume that the data has a known and specific distribution, often a Gaussian distribution. If a

machinelearningmastery.com