

## < 비모수 통계기법 >

① 비모수 통계분석을 사용해야 하는 상황

① 분포 가정 불만족 : ex) 정규분포 가정 X

② 분석기법의 전제조건 불만족 :

③ 표본 수 불만족 : 30개 미만의 표본수 취된 경우

④ 데이터 척도 단수화 : 명목 또는 서열척도로만 측정된 데이터

위와 같은 상황에서 모수 통계기법 대신 사용할 수 있다.

그렇지 않은 경우 1종 오류의 값이 커지거나, 분석 결과 자체에 대한 신뢰성이 떨어지거나 하는 문제점이 발생할 수 있다.

## 비모수 통계분석 유형

크게 적합도검정, 동일성검정, 그리고 상관성검정을 위한 비모수 통계기법으로 분석할 수 있다.

그리고 표본의 모집단의 분포는 단순히 연속성이나 대칭성 정도만 충족하면 된다.

### 적합도검정

#### RUN 검정

• 표본의 배열이 무작위로 구성되어있는지 검정

• ex)  $H_0$  : 귀무가설

멤버십 소지 고객과 비소지 고객 방문은 무작위로 이루어졌다  $\Leftrightarrow H_1$  : 대립가설

#### Kolmogorov-Smirnov 검정

• 표본의 분포와 가정한 분포를 따르지 않는지 검정

• ex)  $H_0$  : T여행객 고객의 구매도점수는 정규분포를 따른다.

$\Leftrightarrow H_1$  : (기존 고객 선호도 결과 정규성을 따르지 않는다)

### 동일성검정

#### Wilcoxon 부호-순위 검정

• 쌍체표본 t-검정과 유사

쌍체표본의 부호와 서열로 분포를 비교하는 검정

•  $H_0$  : ex) 기존 물류 입고량과 신규 입고량을 통한 평균 배송시간의 차이가 없다.

#### Mann-Whitney U 검정

• 독립표본 t-검정과 유사

두 개의 독립표본이 같은 분포를 따르는지 검정

• ex)  $H_0$  : 공장 A와 공장 B의 생산효율은 동일하다

동일하다

#### Kruskal-Wallis H 검정

• 일원 분산분석과 유사

• 2점 이상의 독립표본 집단의 분포를 비교하는 검정

• ex)  $H_0$  : 3개의 공장의 연간 생산량은 동일하다.

## Friedman 검정

• 일원 분산분석과 유사

• 3집단 이상의 쌍체표본 집단의 분포를 비교하는 검정

$H_0$  : 해당 (처치) 별 고객의 선호도 평가에 큰 차이가 없다.

### 상관성검정

#### 스피어만 서열상관계수

• 서열척도 또는 비등척도 자료로 평가대상 간 일관성을 검정

$H_0$  : A 쇼핑객들의 1회 평균 구매액액수와 방문빈도 순위의 연관성이 없다.

• 특히 순서형 변수의 상관관계 분석에 사용

#### Kendall's 서열상관 분석

• 서열척도 또는 비등척도 자료로 평가대상 간 일관성을 검정

$H_0$  : A 홈쇼핑의 의류 브랜드 가격서열과 판매량 서열의 연관성이 없다.

## <RUN 검정>

런(RUN) 검정이란 관측치들이 무작위로 추출되었는지의 여부를 판단하는 비모수 통계분석 기법이다.

RUN 검정에 사용하는 데이터는 연속적으로 이루어진 동일한 관측치의 집합이다.

예를 들어 안정 착용 0 또는 착용 1 이라 할 때

무작위 10명을 뽑아 연달아 생각해 경우  $n$  은 구한다

111\00\1\000\1  $n=5$  경우  
 $n_1=5$ 명  
 $n_0=5$ 명

런 검정 수행방법은 평균과 표준편차를 이용하는 방법과 순위수를 이용하는 방법이 있다.

## o RUN 검정분석과정(순위수 이용)

from statsmodels.stats.runs import Runs

Runs(data).runs\_test()

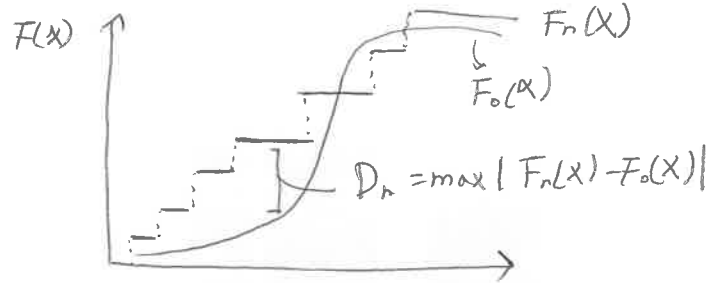
이진데이터만 적용 가능)

결과: (Z값, p값 (양측검정))  
 ↳ 양측검정은 P/2 해주어야 함  
 관찰량

## <Kolmogorov - Smirnov 검정>

위검정은 표본의 분포가 가정한 분포와 적합한지 검정하는 것으로 관측된 자료의 크기를 나열한 후 각 관측치들의 누적 확률을 가정한 분포의 누적 확률과 비교하여,

두 누적 확률 차이의 최대값이 임계치보다 작을 경우 표본은 가정한 분포의 모집단에서 추출되었다고 할 수 있다.



## o Kolmogorov - Smirnov 검정 분석 과정

from statsmodels.stats.diagnostic import kstest\_norm  
 소문자  
 kstest\_norm(data, dist = 'norm')  
 정규분포  
 'exp' ↳ 지수분포

결과값: (검정통계량 Z값, p값)