

Chapter04

통계분석 II

I비모수 검정

M T W T F S S

FASTCAMPUS
ONLINE

금융공학/퀀트 I

강사. 장순용

I 키포인트

- 부호검정 (한 개의 모집단 or 두 개의 모집단).
- Mann-Whitney U 검정.
- Wilcoxon 검정.
- Shapiro-Wilk 검정.

I 비모수 검정 (Non-parametric hypothesis test)

- 다음과 같은 경우 비모수 통계에 해당한다.
 - ⇒ 모집단의 확률분포에 대한 전제가 없다.
 - ⇒ 표본으로 계산한 통계량이 모집단의 모수와 통계적 관계가 없다.
- (~표본의 크기가 작다)

I 부호 검정 (Sign test): 모집단 한 개

- 부호 검정은 ‘+’라는 사건이 ‘-’라는 사건보다 더 많은지를 검정한다.

예). 어느 회사의 월간 수익률이 ‘+’인 경우가 ‘-’인 경우보다 많은지 검정.

- 부호는 ‘+’ 또는 ‘-’ 이므로 이분법적인 상황이다.

⇒ 성공확률 (‘+’의 확률) $p=0.5$ 인 이항확률분포 $\sim \text{Binom}(n, p)$.

- 그러므로 이항분포를 사용하여 다음 가설검정을 실행한다.

I 부호 검정 (Sign test): 모집단 한 개

문제: 시행의 횟수가 6인 경우 유의수준이 5%라면 표본에서 ‘+’의 개수가 몇 개이상 이어야 부호 검정의 귀무가설을 기각할 수 있는가?

I 부호 검정 (Sign test): 모집단 한 개

문제: 시행의 횟수가 6인 경우 유의수준이 5%라면 표본에서 ‘+’의 개수가 몇 개이상 이어야 부호 검정의 귀무가설을 기각할 수 있는가?
(임계치)

$p=0.5$ 를 전제하고 이항확률을 계산해 본다. ‘+’의 횟수를 Q 로 표기한다.

$$P(Q \geq 6) = P(Q = 6) = 0.0156$$

$$P(Q \geq 5) = P(Q = 5) + P(Q = 6) = 0.09375 + 0.0156 = 0.109375$$

유의수준 5%에 해당하는 임계치는 5와 6 사이에 있다. 그러므로

I 부호 검정 (Sign test): 모집단 두 개

- 짝 지어진 두개의 관측치 사이에 양의 편향성이 있는지 검정하고자 한다.

예). 동일한 환자 대상으로 약 투여 전과 후 사이의 차이에 대한 검정.

예). 동일한 환자 대상으로 치료약 X 와 Y 사이의 차이에 대한 검정.

- 두개의 확률변수 X 와 Y 가 있다고 가정하고, $p = P(X > Y)$ 라고 정의하고, 다음 가설을 검정한다.

I 부호 검정 (Sign test): 모집단 두 개

문제: 신약 X 와 기존의 약 Y 를 비교하여 효과 지속시간에 대한 검정을 하려고 한다. 효과 지속시간의 차이 $X-Y$ 의 부호 검정을 환자 6명 대상 유의수준 5% 로 적용하고자 한다. 귀무가설의 유지 또는 기각 조건을 구하여라.

I 부호 검정 (Sign test): 모집단 두 개

문제: 신약 X 와 기존의 약 Y 를 비교하여 효과 지속시간에 대한 검정을 하려고 한다. 효과 지속시간의 차이 $X-Y$ 의 부호 검정을 환자 6명 대상 유의수준 5%로 적용하고자 한다. 귀무가설의 유지 또는 기각 조건을 구하여라.

$p=0.5$ 를 전제하고 이항확률을 계산해 본다. 이제 Q 는 $x_i - y_i$ 의 부호가 '+'인 횟수를 나타낸다. 이전 문제에서 구했듯이 $P(Q \geq 6) = 0.0156$ 이고 $P(Q \geq 5) = 0.109375$ 이다. 그러므로 '+'가 5개까지는 귀무가설을 유지하고 '+'가 6개이면 귀무가설을 기각한다. 모집단 한 개인 경우와는 '+' 부호의 의미에 대한 차이점이 있다.

I Mann Whitney U 검정

- 두 집단 (A 와 B) 간의 차이 여부를 검정하고자 한다.
- 다음의 경우 적용한다:
 - ⇒ 자료의 수치가 서열척도 이다 (평균, 분산 등을 직접 계산 못 함).
 - ⇒ 표본의 크기가 30미만이어서 정규성을 충족시키지 못하는 경우.
- 귀무가설과 대립가설은 다음과 같다.
 - 귀무가설 H_0 : A의 순위 평균은 B의 것과 같다.

I Mann Whitney U 검정

- 순위는 두 개의 표본을 통합하여 구한다.

예). 남학생의 시험점수: 100, 80

여학생의 시험점수: 95, 75, 68



통합된 시험점수 {100, 95, 80, 75, 68} 를 순위로 변환 {1, 2, 3, 4, 5}



남학생의 시험점수 (순위척도): 1, 3

여학생의 시험점수 (순위척도): 2, 4, 5

남/여학생의 시험점수의 순위에서 차이가 있는가? \Rightarrow Mann Whitney

I Wilcoxon 검정

- Wilcoxon 검정을 부호화된 순위 검정 (signed rank test)라고도 부른다.
- 순위와 ‘+’ or ‘-’ 부호를 동시에 고려한다.
- 짝이 지어진 두 집단 (A 와 B) 간의 차이 여부를 검정하고자 한다.

귀무가설 H_0 : A와 B사이에는 차이가 없다.

대립가설 H_1 : A와 B사이에는 차이가 있다.

I Wilcoxon 검정

- 부호화된 순위를 계산하는 방법을 다음 예를 통해서 알아본다.

예).

i	x_i	y_i	$x_i - y_i$	부호	$ x_i - y_i $	$ x_i - y_i $ 의 순위	부호화된 순위
#1	5.5	4.5	1	+	1	1	+ 1
#2	1	17	-16	-	16	6	-6
#3	10	6	4	+	4	5	+ 5
#4	16	14	2	+	2	2.5	+ 2.5
#5	13	11	2	+	2	2.5	+ 2.5
#6	8	5	3	+	3	4	+ 4
#7	9	9	0	없음	0	제외	제외

⇒ 이 경우 부호화된 순위의 합은 9이다. 이것이 0과 의미있는 차이가 있는지 검정한다.

I Shapiro-Wilk 검정

- 표본의 ‘정상성’ (정규분포를 따름)을 검정하고자 한다.
- QQ plot과 같은 시각화 방법보다 선호된다.
- 귀무가설과 대립가설은 다음과 같다.

귀무가설 H_0 : 표본은 정규분포를 따르는 모집단에서 추출되었다.

대립가설 H_1 : 표본은 정규분포를 따르지 않는 모집단에서 추출되었다.

I 끝.

감사합니다.

