

비모수 검정

Team.6 김진형

김규진

비모수 검정이란?

모집단에 대한 추론

모수 통계학

비모수 통계학

모수 검정

비모수 검정

~~~iid(identically distributed + independent)~~

T분포 / Z분포 / F분포

부호 / 순위 관련 분포

# 비모수 검정이란?

## 비모수 검정(Nonparametric Statistics)이란?

- 순위(RANK)와 부호(SIGN)에 기초한 통계 방법
- 모집단의 모수가 필요하지 않음.
- 모수를 다루지 않지만, 기존의 모집단에 대한 추론을 하는 통계학 기본 틀 유지

### POINT 1.

그렇다면 모수에 대한 가정이 전혀 없는가? **NO!**

-> 모집단의 특정 분포를 가정하지 않을 뿐! 모집단의 분포는 존재하며, ~iid의 가정사항은 존재!

### POINT 2.

중앙값 or 중앙값 차이에 대한 검정 <- “평균(**모수**)”에 대한 것으로 바꾸는 것은 가능한가? **YES!**

-> 모집단의 분포가 대칭이라는 “추가적 가정” 필요!

# 비모수 검정이란?

## 비모수 검정의 “장점”

- 대부분의 비모수적 방법 = 가정 최소화
  - > 잘못 적용될 가능성 적음.
- 순위(Rank)와 부호(Sign)을 이용하므로 직관적
  - > 통계지식 없이 이해하기 쉬움
- 모수적 방법에 비해 크게 효율이 떨어지지 않으면서 안전하다

## 비모수 검정의 “단점”

- 분포함수에 대한 가정이 대부분 만족되는 모수검정에 비해 검정력이 떨어진다.
- 대부분 간단한 계산을 요구하지만, 경우에 따라 매우 지루하고 계산이 필요하기도 하다.

# 비모수 검정이란?

## 비모수 검정을 쓰기에 적합한 조건?

- 1) 검정하고자 하는 가설에 “모수”(평균, 분산, 표준편차)가 없는 경우
- 2) 검정하고자 하는 가설에 “분포에 대한 가설”이 없는 경우
- 3) 적합한 모수적 방법을 사용할 만한 데이터가 없는 경우  
-> 즉, 실험(조사)에 따라서는 반응변수 값을
  - (1) 수치로 측정하기 어렵다
  - (2) 수치는 큰 의미가 없고, 그 상대적 크기(Rank, Sign)가 의미가 있을 때
- 4) 모수적 방법을 사용하기 위한 가정이 위반되는 경우

=>,  $n$ 이 작아서 대표본근사(중심극한정리)가 적용되지 않는다 = “정규분포에 대한 가정 불충족”

중심 경향을 나타내는 척도 : Mean(평균) Median(중앙 값)

모수적 방법 : 평균 이용 (중앙 값 이용 가능)

비모수적 방법 : 중앙 값 이용 (Nominal Or Ordinal Data)

$P(X \geq M) = P(X \leq M) = 1/2$ 인  $M$ (중앙값)

# 비모수란 무엇인가?

예시로 이해해보자!

| (n주) | 1  | 2  | 3  | 4  | 5  | 6  | 7  | 8   | 9   | 10  | 11  | 12  |
|------|----|----|----|----|----|----|----|-----|-----|-----|-----|-----|
| 판매량  | 45 | 32 | 39 | 29 | 64 | 55 | 38 | 212 | 187 | 124 | 320 | 188 |

20년 1월부터 3월(12주)까지 주단위 별 핸드폰 판매량은 위와 같다. 이 때, 작년 19년 1월부터 3월(12주)까지의 동일한 기간 동안 핸드폰 판매량 평균이 100이라면, 올해 핸드폰 판매량이 작년 판매량보다 더 높다고 말할 수 있는가? (단,유의수준  $\alpha=0.05$ )

$$H_0 : \mu = 100 \quad \text{versus} \quad H_a : \mu > 100$$

• Z검정 or T검정을 사용한다.

->2가지 가정이 필요하다.

1.정규분포를 따른다 -> [Z분포,T분포 사용]

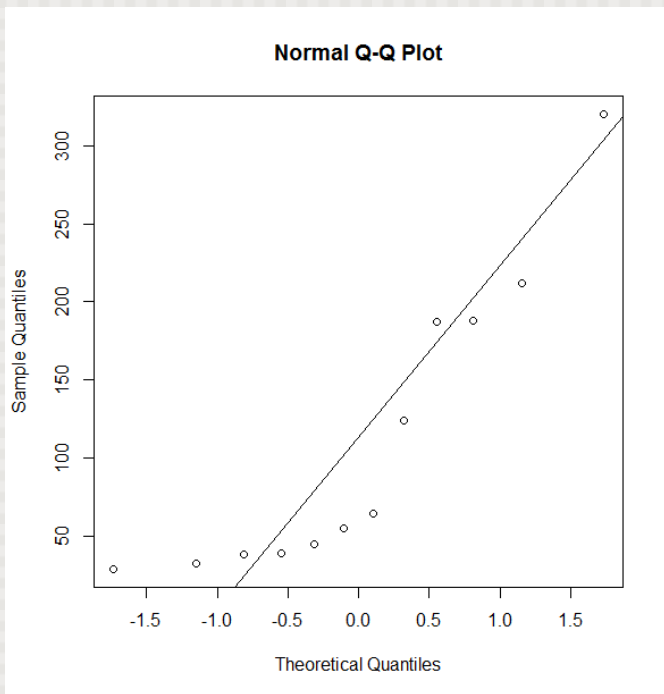
or

2.표본의 크기가 30이상이다. (대체로,  $n > 30$ 을 사용한다) ->대표본 근사[Z분포 사용]

# 비모수란 무엇인가?

위 자료는 앞의 2번 가정[ $n > 30$ ]을 충족시키지 못하므로, 1번 가정[정규분포 가정]을 확인한다.

## 1)qqnorm+qqline



```
> sales <- c(45,32,39,29,64,55,38,212,187,124,320,188)
> qqnorm(sales); qqline(sales)
```

->qqline에서 벗어나므로, 정규성 가정을 만족하지 않는다.

# 비모수란 무엇인가?

위 자료는 앞의 2번 가정[ $n > 30$ ]을 충족시키지 못하므로, 1번 가정[정규분포 가정]을 확인한다.

2)shapiro.test()

```
> sales <- c(45,32,39,29,64,55,38,212,187,124,320,188)
> qqnorm(sales); qqline(sales)
> shapiro.test(sales)
```

shapiro-wilk normality test

```
data: sales
W = 0.82349, p-value = 0.01755
```

P-value가 유의수준(0.05)보다 작으므로 기각  
-> 정규성 가정을 만족하지 않는다.

- 결론

분포에 대한 가정이 성립되지 않기 때문에, 비모수 검정을 사용한다.



# 부호 검정(Sign Test) wilcoxon signed-ranks test

## 부호 검정(Sign Test) – “ONE” sample

|     | 1    | 2    | 3    | 4    | 5    | 6    | 7   | 8   | 9    | 10   |
|-----|------|------|------|------|------|------|-----|-----|------|------|
| 표본값 | 11.2 | 11.0 | 12.0 | 13.0 | 12.5 | 10.5 | 5.8 | 6.2 | 14.0 | 12.2 |
| 부호  | +    | +    | +    | +    | +    | +    | -   | -   | +    | +    |

$H_0$  : 모집단의 중앙값 = 10    versus     $H_a$  : 모집단의 중앙값  $\neq$  10

- 중앙값보다 클 확률  $p = 0.5$ 이다.
- $X$ =중앙값보다 큰 표본의 개수

중앙값보다 큰 표본의 개수가 8이므로,  $\Pr(X \geq 8) = \sum_{i=8}^{10} \binom{n}{i} \left(\frac{1}{2}\right)^i \left(1 - \frac{1}{2}\right)^{n-i} = 0.0547$

-> 단측 검정 = P-value = 0.0547 -> **기각 x** (유의수준 0.05)

-> 양측 검정 = P-value = 0.109 -> **기각 x** (유의수준 0.05)

## 부호 검정(Sign Test) wilcoxon signed-ranks test

부호 검정(Sign Test) – “*ONE*” sample – “*in R*”

```
> x=c(11.2,11.0,12.0,13.0,12.5,10.5,5.8,6.2,14.0,12.2)
```

```
> sign.test=function(median,y){  
+   z=sort(y)  
+   n=length(z)  
+   b=sum(z>median)  
+   pbinom(b-1,n,0.5,lower.tail=FALSE)  
+ }
```

```
> sign.test(10,x)  
[1] 0.0546875
```

# 부호 검정(Sign Test) wilcoxon signed-ranks test

## 부호 검정(Sign Test) – “ONE” sample – “in R”

- 비모수검정 (Sign Test)

```
> sign.test(10,x)
[1] 0.0546875
```

- 모수검정 (T Test)

```
> t.test(x,alternative="two.sided")

one sample t-test

data:  x
t = 12.489, df = 9, p-value = 5.476e-07
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 8.876508 12.803492
sample estimates:
mean of x
 10.84
```

sign Test와 T Test의 “P-Value” 큰 차이 = “기각여부” 차이 남.

# 부호 검정(Sign Test) wilcoxon signed-ranks test

## 부호 검정(Sign Test) – “*PAIRED*” sample

|       | 1   | 2   | 3   | 4   | 5   | 6   | 7   | 8   |
|-------|-----|-----|-----|-----|-----|-----|-----|-----|
| 학습 전  | 750 | 850 | 600 | 550 | 560 | 640 | 800 | 640 |
| 학습 후  | 850 | 840 | 625 | 535 | 620 | 670 | 820 | 680 |
| 성적 차이 | 100 | -10 | 25  | -15 | 60  | 30  | 20  | 40  |
| 부호    | +   | -   | +   | -   | +   | +   | +   | +   |

Ho : 모집단의 중앙값 = 0    versus    Ha : 모집단의 중앙값 > 0

• 중앙값보다 클 확률  $p = 0.5$ 이다.

•  $X$ =중앙값보다 큰 표본의 개수

$$\Pr(X \geq 6) = \sum_{i=6}^8 \binom{n}{i} \left(\frac{1}{2}\right)^i \left(1 - \frac{1}{2}\right)^{8-i} = 0.145$$

-> 단측 검정 = P-value = 0.145 -> **기각 x**(유의수준 0.05)

# 부호 검정(Sign Test) wilcoxon signed-ranks test

## 부호 검정(Sign Test) – “*PAIRED*” sample – “*in R*”

```
> a=c(750,850,600,550,560,640,800,640)
> b=c(850,840,625,535,620,670,820,680)
> d=b-a
> T=length(d[d>0])
```

```
> binom.test(T,length(d[d!=0]),alternative="greater")
```

Exact binomial test

data: T and length(d[d != 0])

number of successes = 6, number of trials = 8, p-value =  
0.1445

alternative hypothesis: true probability of success is greater than 0.5

95 percent confidence interval:

0.4003106 1.0000000

sample estimates:

probability of success

0.75

# 부호 검정(Sign Test) wilcoxon signed-ranks test

## 부호 검정(Sign Test) – “*PAIRED*” sample – “*in R*”

- 비모수검정 (Sign Test)

```
> binom.test(T,length(d[d!=0]),alternative="greater")

Exact binomial test

data:  T and length(d[d != 0])
number of successes = 6, number of trials = 8, p-value = 
0.1445
alternative hypothesis: true probability of success is greater than 0.5
95 percent confidence interval:
 0.4003106 1.0000000
sample estimates:
probability of success
0.75
```

- 모수검정 (T Test)

```
> t.test(a,b,alternative="two.sided",paired=T)

Paired t-test

data:  a and b
t = -2.3821, df = 7, p-value = 0.04873
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -62.2706091 -0.2293909
sample estimates:
mean of the differences
-31.25
```

sign Test와 T Test의 “P-Value” 큰 차이 = “기각여부” 차이 남.

# 윌콕슨 검정(Wilcoxon Signed-Ranks Test)

## 부호 검정의 한계점

부호(Sign)만으로 검정하므로, 표본 값이 “중앙값(M or 0)으로부터 얼마나 벗어나 있는가”는 고려하지 못함!

“Rank”를 추가적으로 고려함으로써 한계 극복!



부호검정(Sign Test) + 검정력 증가 = 부호순위검정(Signed rank Test)

## #POINT

- Sign = 차이의 “방향”(direction)
- Rank = 차이의 “양”(amount)

# 윌콕슨 검정(Wilcoxon Signed-Ranks Test)

## 윌콕슨 검정(Wilcoxon Signed\_ranks Test)

$H_0 : \text{중앙값} = M$     versus     $H_a : \text{중앙값} \neq M$

1) 표본에서  $M$ 을 뺀 차이의 절댓값에 대해 순위를 매긴다

2) 0보다 작으면  $-$  크면  $+$ 를 순위에 부여한다

3)  $+$  순위의 합  $T^+$ 를 구한다.

4) 이 값이 너무 크면 귀무가설을 기각한다.



# 윌콕슨 검정(Wilcoxon Signed-Ranks Test)

## 윌콕슨 검정(Wilcoxon Signed\_ranks Test)

$H_0$  : 중앙값 =  $M$  versus  $H_a$  : 중앙값  $\neq M$

1) 표본에서  $M$ 을 뺀 차이의 절댓값에 대해 순위를 매긴다

2) 0보다 작으면  $-$  크면  $+$ 를 순위에 부여한다

3) "+순위의 합  $T^+$ "를 구한다.

4) 이 값이 너무 크면 귀무가설( $H_0$ )을 기각한다.

Cf) 귀무가설 기각 조건 = 하한 임계값( $T_L$ )과 상한 임계값( $T_U$ )보다 밖에 있는 경우.

-> 기존의 검정통계량 기각 조건과 동일.

## 윌콕슨 부호순위 검정 통계량(T)

$$\text{평균} = \frac{n(n+1)}{4}$$

$$\text{분산} = \frac{n(n+1)(2n+1)}{24}$$

$$T \sim N\left(\frac{n(n+1)}{4}, \frac{n(n+1)(2n+1)}{24}\right)$$

# 윌콕슨 검정(Wilcoxon Signed-Ranks Test)

윌콕슨 부호순위 검정 임계값

$n = ?$ ,  $T^+ = ?$

$T_U = ?$

$T_L = ?$

| n  | a=0.05 |       | a=0.10 |       |
|----|--------|-------|--------|-------|
|    | $T_L$  | $T_U$ | $T_L$  | $T_U$ |
| 6  | 1      | 20    | 2      | 19    |
| 7  | 2      | 26    | 4      | 24    |
| 8  | 4      | 32    | 6      | 30    |
| 9  | 6      | 39    | 8      | 37    |
| 10 | 8      | 47    | 11     | 44    |
| 11 | 11     | 55    | 14     | 52    |
| 12 | 14     | 64    | 17     | 61    |
| 13 | 17     | 74    | 21     | 70    |
| 14 | 21     | 84    | 26     | 79    |
| 15 | 25     | 95    | 30     | 90    |
| 16 | 30     | 106   | 36     | 100   |
| 17 | 35     | 118   | 41     | 112   |
| 18 | 40     | 131   | 47     | 124   |
| 19 | 46     | 144   | 54     | 136   |
| 20 | 52     | 158   | 60     | 150   |
| 21 | 59     | 172   | 68     | 163   |
| 22 | 66     | 187   | 75     | 178   |
| 23 | 73     | 203   | 83     | 193   |
| 24 | 81     | 219   | 92     | 208   |
| 25 | 90     | 235   | 101    | 224   |
| 26 | 98     | 253   | 110    | 241   |
| 27 | 107    | 271   | 120    | 258   |
| 28 | 117    | 289   | 130    | 276   |
| 29 | 127    | 308   | 141    | 294   |
| 30 | 137    | 328   | 152    | 313   |

# 윌콕슨 검정 – “ONE sample”

|         | 1    | 2    | 3    | 4    | 5    | 6    | 7    | 8    | 9    | 10   |
|---------|------|------|------|------|------|------|------|------|------|------|
| 표본값     | 11.2 | 11.0 | 12.0 | 13.0 | 12.5 | 10.5 | 5.8  | 6.2  | 14.0 | 12.2 |
| 중앙값과의 차 | 1.2  | 1.0  | 2.0  | 3.0  | 2.5  | 0.5  | -4.2 | -3.8 | 4.0  | 2.2  |
| 순위      | 3    | 2    | 4    | 7    | 6    | 1    | 10   | 8    | 9    | 5    |
| 부호순위    | 3    | 2    | 4    | 7    | 6    | 1    | -10  | -8   | 9    | 5    |

$T_U, T_L$

[ $H_0$  : 모집단의 중앙값 = 10 versus  $H_a$  : 모집단의 중앙값  $\neq$  10](유의수준 = 0.05)

$n=10, T^+ = 37$

상한 임계값  $T_U = 44$

하한 임계값  $T_L = 11$



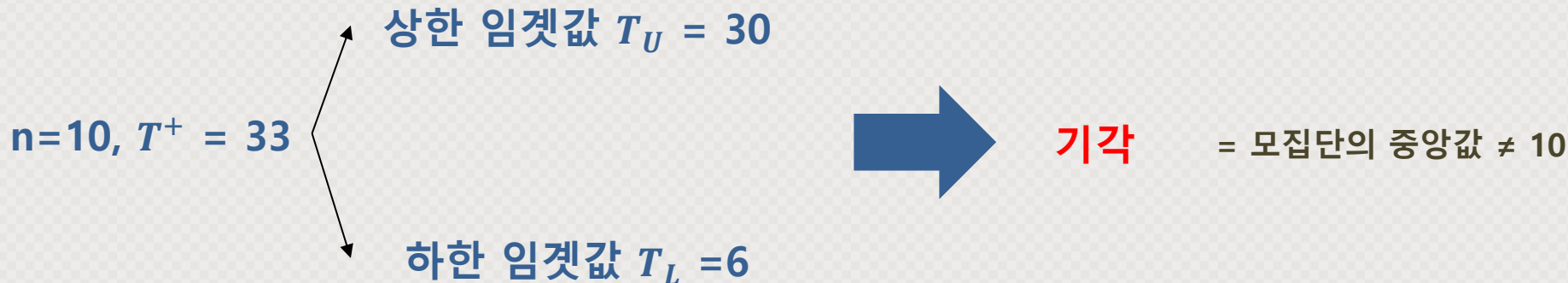
기각X

= 모집단의 중앙값 = 10

# 윌콕슨 검정 – “*PAIRED sample*”

|         | 1   | 2   | 3  | 4   | 5  | 6  | 7  | 8  |
|---------|-----|-----|----|-----|----|----|----|----|
| 성적 차이   | 100 | -10 | 25 | -15 | 60 | 30 | 20 | 40 |
| $ D_i $ | 100 | 10  | 25 | 15  | 60 | 30 | 20 | 40 |
| 순위      | 8   | 1   | 4  | 2   | 7  | 5  | 3  | 6  |
| 부호순위    | 8   | -1  | 4  | -2  | 7  | 5  | 3  | 6  |

$T_U, T_L$       [Ho : 모집단의 중앙값 = 10    versus    Ha : 모집단의 중앙값  $\neq$  10] (유의수준 = 0.05)



# Permutation Test란...?

## Permutation Test

T-test 등의 일반적인 통계 검정을 수행할 만큼 샘플의 수가 크지 않은 경우에 사용할 수 있는 검정 방법이다.

주어진 샘플을 무작위로 추출하여 인공적으로 샘플 숫자를 늘림으로써 전체 모수를 통계 검정이 가능한 크기만큼 키운 다음, 주어진 샘플의 통계 값이 전체 모수와 비교하여 얼마나 유의하게 차이 나는지를 검정하는 방법이다.

M+N개의 실험단위 중에서 각 그룹에 M, N개씩 뽑아서 할당한다고 했을 때, 두 집단의 평균이나 중위수에는 크게 차이가 없을 것이기 때문에 이런 방법을 사용할 수 있다.

# Permutation Test 예제

|          |     |     |     |     |     |     |     |     |     |     |  |
|----------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|--|
| <b>A</b> | 156 | 183 | 120 | 113 | 138 | 145 | 142 |     |     |     |  |
| <b>B</b> | 109 | 107 | 119 | 162 | 121 | 123 | 76  | 111 | 130 | 115 |  |

A그룹의 평균 성장률이 B그룹의 평균 성장률 보다 높은가?

$H_0: \mu_A - \mu_B = 0$

vs

$H_a: \mu_A - \mu_B > 0$

P - value 의 값이 0.0209이므로  
귀무가설을 기각하기 충분함!

```
> A <- c(156,183,120,113,138,145,142)
> B <- c(109,107,119,162,121,123,76,111,130,115)
>
> R <- 9999                                     #Permuted test sample의 size를 지정한다.
> all <- c(A,B)                                #A와 B를 하나의 벡터로 합쳐준다.
> k <- 1:length(all)                           #합친 벡터의 인덱스를 설정해준다.
> reps <- numeric(R)                          #Permuted test의 통계량을 담을 객체를 생성한다.
> ts <- mean(A) - mean(B)                     #검정 통계량의 값을 계산해준다.
> ts
[1] 25.12857
>
> for (i in 1:R) {
+   # 크기 A만큼의 sample을 뽑아 A에 저장한다.
+   m <- sample(k, size=length(A), replace=FALSE)
+   a1 <- all[m]                               #새롭게 뽑은 A를 a1에 저장한다
+   b1 <- all[-m]                             #새롭게 뽑은 B를 b1에 저장한다
+   reps[i] <- mean(a1) - mean(b1)            #i번째 순열검정 통계량을 계산한다.
+ }
> pvalue <- mean(c(ts, reps) >= ts)           #p-value를 계산
> pvalue
[1] 0.0191
```

# *Wilcoxon Rank Sum Test*

## Wilcoxon Rank Sum Test vs Wilcoxon Singed Rank Test

앞서 배운 Wilcoxon Signed Rank Test와 Wilcoxon Rank Sum Test의 차이는 뭘까?

기본적으로,

데이터가 쌍을 이루는 경우는 Singed Rank Test를 사용해야하는 반면,

데이터가 페어링을 이루지 않는 경우 Rank Sum Test를 사용해야한다.

# Wilcoxon Rank Sum Test

## Wilcoxon Signed Rank Test vs Wilcoxon Rank Sum Test

### Wilcoxon Signed Rank Sum Test

- 모집단이 쌍을 이루어야 한다. 예를 들어, 동일한 그룹의 사람들이 서로 다른 두 가지 경우에 테스트되고 각각의 영향에 대해 측정된 후 비교해야 한다.

### Wilcoxon Rank Sum Test

- 모집단이 쌍을 이루지 않아도 된다. 독립 모집단에서 표본을 추출하는 것이다. 예를 들어, 시험지 1이 시험지 2보다 어려운지 테스트하고 싶을 때, 독립적인 두 그룹의 학생이 있고, 그 그룹의 크기 또한 같지 않아도 된다.



# Wilcoxon Rank Sum Test 예제

|          |     |     |     |     |     |     |     |     |     |     |  |
|----------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|--|
| <i>A</i> | 156 | 183 | 120 | 113 | 138 | 145 | 142 |     |     |     |  |
| <i>B</i> | 109 | 107 | 119 | 162 | 121 | 123 | 76  | 111 | 130 | 115 |  |

A그룹의 평균 성장률이 B그룹의 평균 성장률 보다 높은가?

$H_0: \mu_A - \mu_B = 0$

vs

$H_a: \mu_A - \mu_B > 0$

```
> A <- c(156,183,120,113,138,145,142)
> B <- c(109,107,119,162,121,123,76,111,130,115)
> library(exactRankTests)
> wilcox.exact(A, B, paired=FALSE, alternative="greater")
```

Exact wilcoxon rank sum test

```
data: A and B
W = 56, p-value = 0.02154
alternative hypothesis: true mu is greater than 0
```

P - value 의 값이 0.0209이므로  
귀무가설을 기각하기 충분함!

## Bootstrap이 뭘까...?

Bootstrap은 복원 추출을 통해 sample의 개수를 늘리는 기법이다.

앞에서 배운 permutation test에서 resampling하는 것과는 조금 차이가 있지만, Permutation test에서 사용한 resampling의 많은 이점을 같이 가지고 있다.

먼저, 이 resampling이란 한 sample에서 반복적으로 sampling을 하는 것인데, 이는 정규 분포를 따른다거나 표본이 크다거나 등의 가정들을 가지지 않아도 된다.

또한, 일반적인 통계 기법들 보다 더 좋은 정확도를 가지며 어떤 통계적인 방법에도 적용 될 수 있다.

## Bootstrap의 기본 Idea

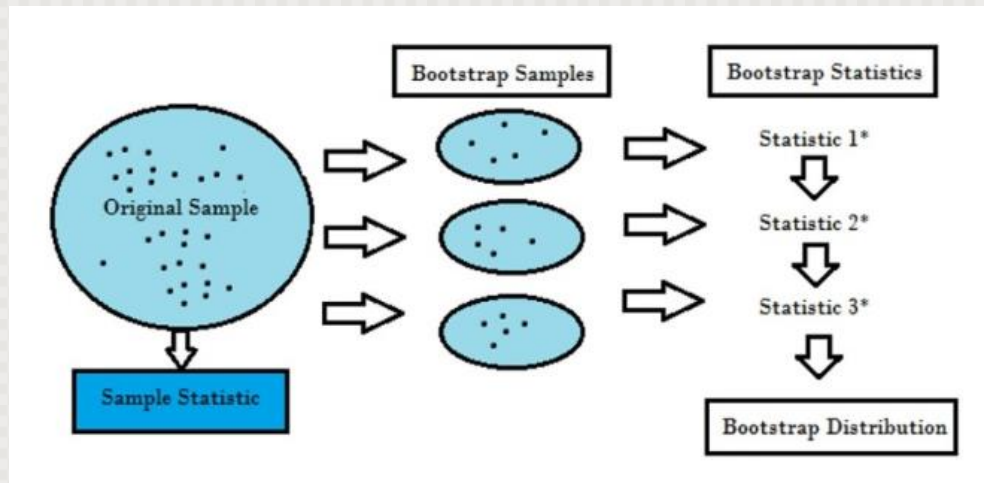
하나의 표본은 그 표본이 추출된 모집단에 대한 정보를 담는다. 따라서, 이 표본에 대해 resampling 방법을 사용함으로 수 많은 표본들을 만들 수 있다.

23.3 26.1 19.0 28.8 29.0

위 크기가 5인 표본을 통해서 우측에 3개 표본을 만들어냈다. 3개가 아니라 몇 백개까지 만들고 싶은 만큼 만들 수 있음!!

```
> x <- c(23.3,26.1,19.0,28.8,29.0)
> x
[1] 23.3 26.1 19.0 28.8 29.0
> bootsample1 <- sample(x, size=length(x), replace=TRUE)
> bootsample1
[1] 29.0 29.0 19.0 23.3 28.8
> bootsample2 <- sample(x, size=length(x), replace=TRUE)
> bootsample2
[1] 26.1 29.0 19.0 26.1 26.1
> bootsample3 <- sample(x, size=length(x), replace=TRUE)
> bootsample3
[1] 23.3 28.8 19.0 19.0 28.8
```

# Bootstrap

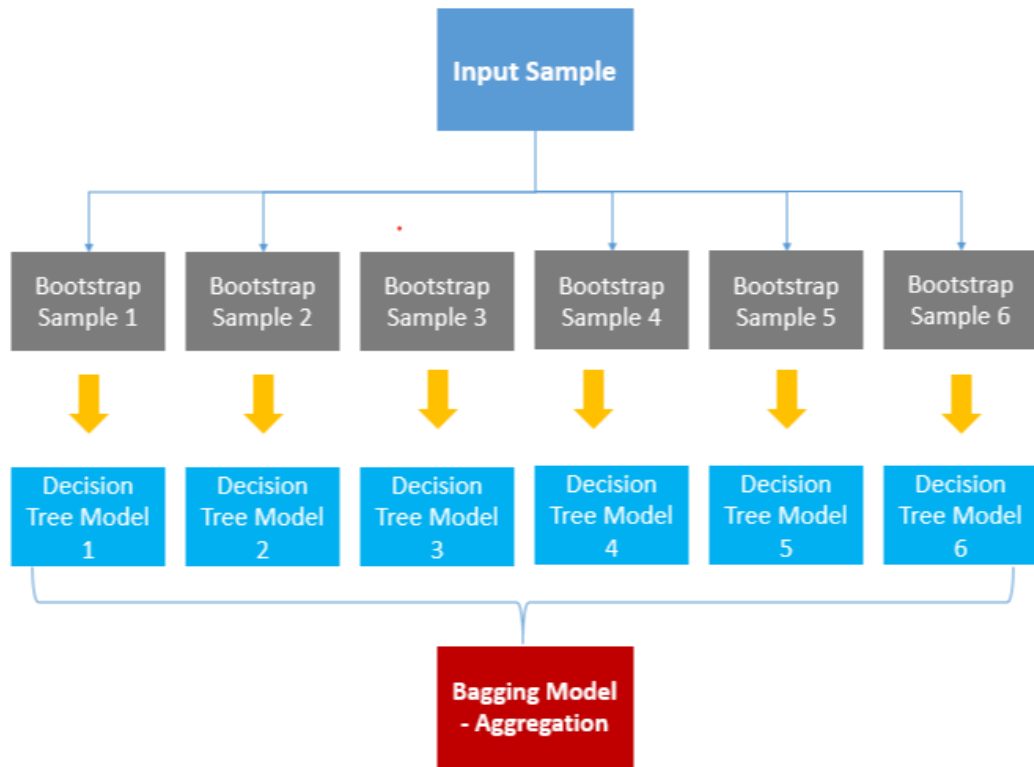


왼쪽의 Original Sample을 하나의 모집단처럼 가정하여 더 작은 크기의 표본집단을 매우 많이 복원추출한다.

이렇게 얻은 Bootstrap Samples로 부터 각각의 통계 값을 평균 처리해준다.

비용과 시간이 많이 드는 데이터 수집을 스스로 해결하고자 할 때, 하나에 표본 집단에 대하여 적용하면 되므로 유용하다. (물론, 이 표본집단의 규모가 어느 정도는 되어야 한다는 조건은 있다.)

# Bootstrap in ML



기계학습에서도 bootstrapping 기법이 사용된다. 우리가 가지고 있는 데이터를 bootstrap을 통해서 여러 개를 만든 후, 그 데이터들을 토대로 각각의 모델을 만들어 마지막으로 모델들을 합쳐서 더 정확한 모델을 만들 수 있다.